INDIAN AGRICULTURAL
RESEARCH INSTITUTE, NEW DELHI.

I.A.R.I. 6.
MGIPC—S4—10 AR—21-6-49—1,000.

PROCEEDINGS
OF THE
ROYAL SOCIETY OF LONDON

SERIES A. MATHEMATICAL AND PHYSICAL SCIENCES

VOL 183

LONDON
Printed and published for the Royal Society
By the Cambridge University Press
Bentley House, N.W.1

18 June 1945

PRINTED IN GREAT BRITAIN BY
WALTER LEWIS, M.A.
AT THE CAMBRIDGE UNIVERSITY PRESS

CONTENTS

SERIES A VOLUME 183

No. A 992. 10 August 1944

PAGE

An investigation of the scattering of high-energy particles from the cyclotron by the photographic method. I. The experimental method. By J. Chadwick, F.R.S., A. N. May, T. G. Pickavance and C. F. Powell 1

Long duration of the Balmer spectrum in excited hydrogen. By Lord Rayleigh, F.R.S. (Plate 1) 26

The thermal decomposition of acetone. By J. R. E. Smith and C. N. Hinshelwood, F.R.S. 33

On the flow of gas through a nozzle with velocities exceeding the speed of sound. By L. Fox and R. V. Southwell, F.R.S. 38

The production of cosmic-ray bursts by mesotrons. By C. B. O. Mohr and G. H. Stafford 54

Applications of the photographic method to problems in nuclear physics. II. The scattering of 8·8 and 13 MeV neutrons by protons. By F. C. Champion and C. F. Powell 64

On the equation of state for solids. By R. Fürth 87

Continuous spectra in flames: the role of atomic oxygen in combustion. By A. G. Gaydon. (Plate 2) 111

No. A 993. 30 November 1944

Relaxation methods applied to engineering problems. X. The graphical representation of stress. By D. N. de G. Allen and R. V. Southwell, F.R.S. . 125

On the theory of point-particles. By H. J. Bhabha, F.R.S. and Harish-Chandra . 134

On the removal of the infinite self-energies of point-particles. By Harish-Chandra 142

The combination of fibrous proteins with acids. II. The adsorption of dye anions. By G. A. Gilbert 167

The transition effect of penetrating showers. By L. Jánossy and G. D. Rochester 181

The barometer effect of penetrating showers. By L. Jánossy and G. D. Rochester 186

Some aspects of the production of mesons and the barometer effect of penetrating showers. By L. Jánossy 190

Statistical thermodynamics of mixtures with zero energies of mixing. By E. A. Guggenheim 203

Statistical thermodynamics of mixtures with non-zero energies of mixing. By E. A. Guggenheim 213

The initiation of long electrical discharges. By C. E. R. Bruce 228

No. A 994. 22 February 1945

PAGE

Anniversary Address by Sir Henry Dale, O.M., G.B.E. 243

Relaxation methods applied to engineering problems. XI. Problems governed by the 'quasi-plane-potential equation'. By D. N. de G. Allen, R. V. Southwell, F.R.S. and Gillian Vaisey 258

Unitary representations of the Lorentz group. By P. A. M. Dirac, F.R.S. . . 284

An investigation of the performance of the Rayleigh disk. By R. A. Scott . . 296

A note on the Hamiltonian theory of quantization. By T. S. Chang 316

The occurrence and properties of molecular vibrations with $V(x) = ax^4$. By R. P. Bell 328

Absorption of radium (B + C) gamma-rays. By J. E. Roberts 338

No. A 995. 18 June 1945

The normal vibrations of bridged X_2Y_6 molecules. By R. P. Bell and H. C. Longuet-Higgins 357

The dissociation energies of CO, N_2, NO and CN. By A. G. Gaydon and W. G. Penney 374

The structure of polyisoprenes. I. The crystal structure of geranylamine hydrochloride. By G. A. Jeffrey 388

The statistical theory of the strength of bundles of threads. I. By H. E. Daniels 405

Stationary electron swarms in electromagnetic fields. By D. Gabor 436

The electrical conductivity of an ionized gas in a magnetic field, with applications to the solar atmosphere and the ionosphere. By T. G. Cowling 453

Index 481

An investigation of the scattering of high-energy particles from the cyclotron by the photographic method

I. The experimental method

By J. Chadwick, F.R.S., A. N. May, T. G. Pickavance and C. F. Powell

(*Received* 14 *September* 1943)

The advantages to be derived from a successful application of the photographic method of detecting nuclear particles to the study of the scattering of the high-energy beams of particles provided by the cyclotron are discussed. The design of a 'camera' for this purpose is described. The most important feature of the method lies in the fact that a single plate, exposed to the scattered particles for a few minutes, contains the information required for a complete examination of the variation of the scattered intensity in the angular range from 10 to 170° with the direction of the primary beam. A very great saving is thus effected in the time the cyclotron is occupied in making such experiments as compared with other methods of detection.

The theory of the method is discussed in some detail, and tests of the performance of the instrument are described which demonstrate its reliability. It is shown that the angular resolving power of the method is of the order of $\pm 2°$, and that its ability to distinguish different groups of scattered and disintegration particles from range distribution curves is superior to that hitherto achieved with counter and other methods of detection.

A. DESCRIPTION OF THE METHOD

1. Introduction

The investigation of the scattering of charged particles by atomic nuclei is still one of the most powerful methods of determining the nature of the nuclear forces, and many experiments have been made during the past few years using artificially accelerated protons and deuterons in addition to the α-particles from the natural radioactive elements. As a result it has become clear that it is important to extend these investigations to the high-energy particles which can be produced only with the cyclotron. Such experiments are attended, however, with considerable difficulties if counter or expansion-chamber methods of detecting the scattered particles are employed.

The technical difficulties arise from the following causes. In order to make experiments in conditions of good geometry, which, as will be shown later, are highly desirable if the finer details of the scattering processes are to be elucidated, it is necessary to define a narrow beam of particles from the cyclotron with a restricted angle of divergence. The currents then available are at present rather small; of the order of 10^{-7} to 10^{-8} A. Secondly, it is desirable to use thin targets in order to be able to study the different scattering processes in conditions where the primary beam is homogeneous in energy. Otherwise, with thick targets, it is difficult to determine the variation of the efficiency of a particular process with the energy of the primary beam and, further, the particles produced in a nuclear reaction with a definite energy release do not appear as a homogeneous group but have a spread in energy corresponding to the variation in the energy of the primary beam in traversing the target. In such conditions it is not possible to resolve different groups and to study them individually. With this second restriction the yield of scattered and disintegration particles becomes inconveniently small.

The low yield of scattered particles is particularly serious in work with primary beams of fast deuterons when counter methods of detection are employed, because they have to be recorded against a high background of impulses produced by the heavy general neutron flux from the cyclotron tank (Pollard & Watson 1940). As a consequence of this background the resolving power of the method for groups of different range is restricted even when using proportional counters in coincidence for the detection of the scattered particles. It has been pointed out by Bower & Burcham (1939) that the detection of the scattered particles by the expansion chamber allows high resolving power without serious restriction of solid angle, and Schultz, Davidson & Ott (1940) have used the method to detect the scattered particles resulting from the bombardment of neon by a beam of deuterons with an energy of 3·8 MeV. Whilst an improved resolving power was thus obtained the method is too laborious to allow a detailed examination of the variation of the scattered intensity with angle to be attempted.

It seemed clear to us that the photographic method of detection offered very great advantages in dealing with this problem. It has been shown (Powell, May, Chadwick & Pickavance 1939, 1940, 1943) that the method is reliable for the determination of the energy and the intensity of homogeneous groups of particles, and for the present purpose it has the great advantage that the particles scattered through different angles can all be recorded at the same time, thus eliminating the necessity of monitoring the primary beam. Since the scattered particles of all energies and at all angles of scattering are recorded simultaneously, the time for which the cyclotron is occupied in making a single exposure is enormously reduced compared with that required for the same investigation using the other methods of detection. Further, the single exposure gives a permanent record of many hundreds of times as many tracks as are required for a satisfactory investigation of the variation of the scattered intensity over the complete angular range from 10 to 170° with the direction of the primary beam. Most of this wealth of experimental material is normally not employed, but any part of it can be used if particular regions of the distribution of the scattered particles show interesting features which it is desirable to examine in detail.

In a previous note (Powell *et al.* 1940) we have given a preliminary account of the results of experiments on the scattering of protons which show that these advantages can be realized, our conclusions in this respect being substantially the same as those of Wilkins & Kuerti (1939, 1940) and Wilkins (1941). In this and subsequent papers we give a detailed account of the methods employed and the results obtained for the scattering, by twelve different elements, of 4 MeV protons and 7 MeV deuterons, produced by the Liverpool cyclotron. The great economy which the method effects in the operation time of the cyclotron is illustrated by the fact that the exposures from which the present results were obtained were all made in two periods each of 3 days. It will be shown that even in work with deuteron beams the proton recoils produced by the neutron background give rise to no serious difficulties, and that the experimental width of homogeneous groups of particles appears to be considerably less than that achieved hitherto in similar experiments using counter methods of detection. We would emphasize that these papers are designed to display the power of the method and to make a rough survey of the field which is thus opened for investigation. They do not constitute a finished investigation, but the results enable us to show the most suitable reactions for a much more detailed study in which experiments will be made over a wide range of energies of the primary particles, and we shall make these experiments when a suitable opportunity occurs.

2. DESCRIPTION OF THE APPARATUS

Wilkins (1941) has described a camera for use with the cyclotron in which sixty-six photographic plates are adjusted radially about a scattering foil, each plate recording particles scattered at a single angle with respect to the primary beam. Sixty-six separate shutters were employed, each of which could be opened and shut during exposure to give a suitable density of tracks on the corresponding plate. We have preferred to use a camera with a single large rectangular plate, and to study the scattering, when possible, from gas 'targets'. The essential features of the method were described in our previous note.

In our design a defined beam from the cyclotron passes down the axis of a tube parallel to, and just above the plane of the rectangular photographic plate. This tube has an interruption through which some of the particles scattered by the gas filling the camera can escape and enter the photographic emulsion at a glancing angle to the surface. The essential features of this arrangement are shown in figure 1. It has the advantage that a single plate contains information on the scattering from 10 to 170°, providing regions at any angle of scattering in this range, in which the density of tracks is convenient for counting purposes.

The stray field of the cyclotron magnet causes a considerable curvature in the path of a beam of ions emerging from the target chamber. The scattering apparatus is therefore placed about 2 ft. from the pole-piece of the magnet, where the stray field is relatively weak, and the beam is brought to it down an evacuated tube, shown in figure 2, which is bolted on to the face of the target chamber. This tube is bent to follow the estimated path of the beam, and has a section of tombac tubing T_1, so that it can be moved while the cyclotron is running, until the spot formed by the impact of the beam emerging from the mica window s_2 on a fluorescent screen has the maximum brightness and is uniformly illuminated. The entrance stop, s_1, is 12 mm. in diameter and limits the angular divergence of the beam leaving s_2, which is 3 mm. in diameter, to 1°. The intermediate stops are designed to prevent particles scattered at the walls of the tube from reaching s_2. The tube is evacuated, through s_1 and the target chamber, by the main pumps.

The beam, 3 mm. in diameter and with a divergence of 1°, emerges from the stop s_2, and enters the tube B (figure 2) which has an internal diameter of 15 mm. This tube can be adjusted by the tombac joint T_2 until the beam emerges centrally from the far end as shown by the illuminated patch which it produces on a fluorescent screen. When these adjustments have been made, the screws controlling the movement of the tombac joints are locked in position so that the position of tube B is fixed.

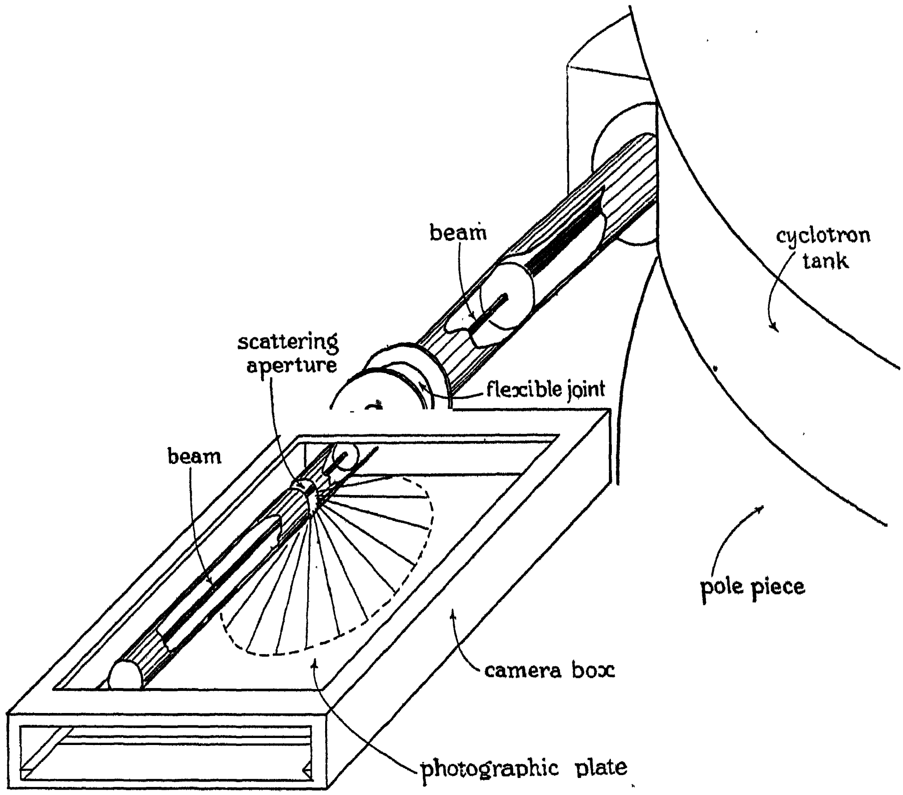


FIGURE 1. Perspective drawing to show disposition of main components of the apparatus.

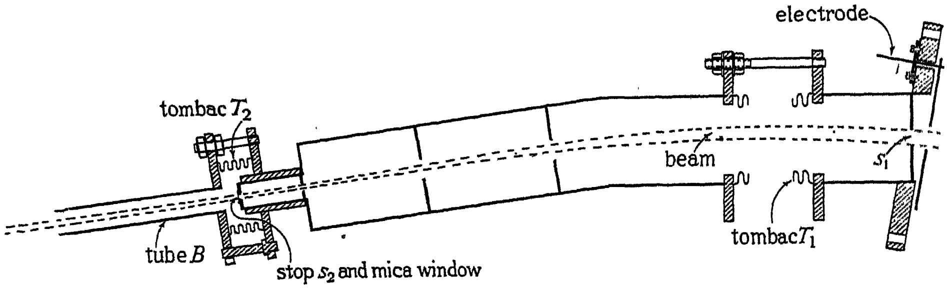


FIGURE 2. Bent tube and slit system for the definition of the beam in the stray field of the magnet.

The scattering apparatus is shown in figure 3. Tube B is a sliding fit on a guide tube carried by the front wall of the camera box. The 'camera' is mounted on levelling screws on a carriage moving on rails, not shown in figure 3. The rails are first adjusted to be parallel to the axis of tube B. The carriage is then placed on the rails and the levelling screws defining the disposition of the camera are adjusted until the axis of the guide tube on the camera coincides with the axis of the tube B. The camera can then be moved so that the tube B enters the guide tube and moves forward until it reaches a position defined by a stop. When this operation has been completed there is a gap of 3 mm., at the point g, figure 3, between the end of the tube B and the opposing edge of the co-axial tube C. The edges of the two tubes are chamfered externally at the gap g to a feather edge at an angle of 5°.

Particles scattered from the beam by the gas filling the camera can emerge through the gap g to enter the emulsion of the plate, 11 × 4 in., which is carried by a plate holder sliding in grooves on the floor of the camera box. The surface of the emulsion is 14·5 mm. below the level of the axis of the tube B and parallel to it. The first stop carried by the tube B, which is 2 mm. in diameter, gives the final definition of the beam. Subsequent stops are clear of the beam and prevent particles scattered by the mica window or by the edges of the stop from reaching the gap g and passing through it. A thin aluminium foil, with an equivalent thickness of 1 cm. of air, covers the gap and prevents light generated by the passage of the beam through the gas from fogging the plate. Copper foils, with an equivalent thickness of 60 and 120 cm. air, can also be rotated into the path of the scattered beam for use in experiments in which it is desired to examine only the high-energy protons arising from disintegration reactions, when a deuteron beam is being used to bombard certain of the light elements. The disposition of these foils is shown in section in figure 3*d*.

An ion collector inside the target chamber and just in front of stop s_1 (figure 2) is used to measure the total beam current and to indicate resonance between the magnetic and high-frequency fields of the cyclotron. Only a small fraction of the total beam current, which was generally of the order of 5×10^{-6} A, reaches the tube B where the beam current is of the order of 10^{-8} A.

It is important that the position of any point on the photographic plate with respect to the mid-point of the gap g shall be known to an accuracy of 0·1 mm. The position of the plate carrier in the camera is defined by a stop, not shown in figure 3, which comes against the front wall of the camera box. The position of the plate in the carrier is determined, after exposure to the beam, by inserting in place of the usual dark cover slide an alternative slide which has been drilled with a rectangular network of holes 0·5 mm. in diameter and 1 cm. apart. This perforated slide is brought to a standard position in the carrier by means of a mark which is brought to coincide with a corresponding mark on the carrier. On exposure to a distant source of light for 1 sec. before development a set of fiducial spots is thus imposed on the plate from which the position of any point on the plate relative to the gap can be determined.

The plate carrier is provided with four strong leaf springs which hold the plate down at the four corners and prevent relative movement of the plate and its carrier.

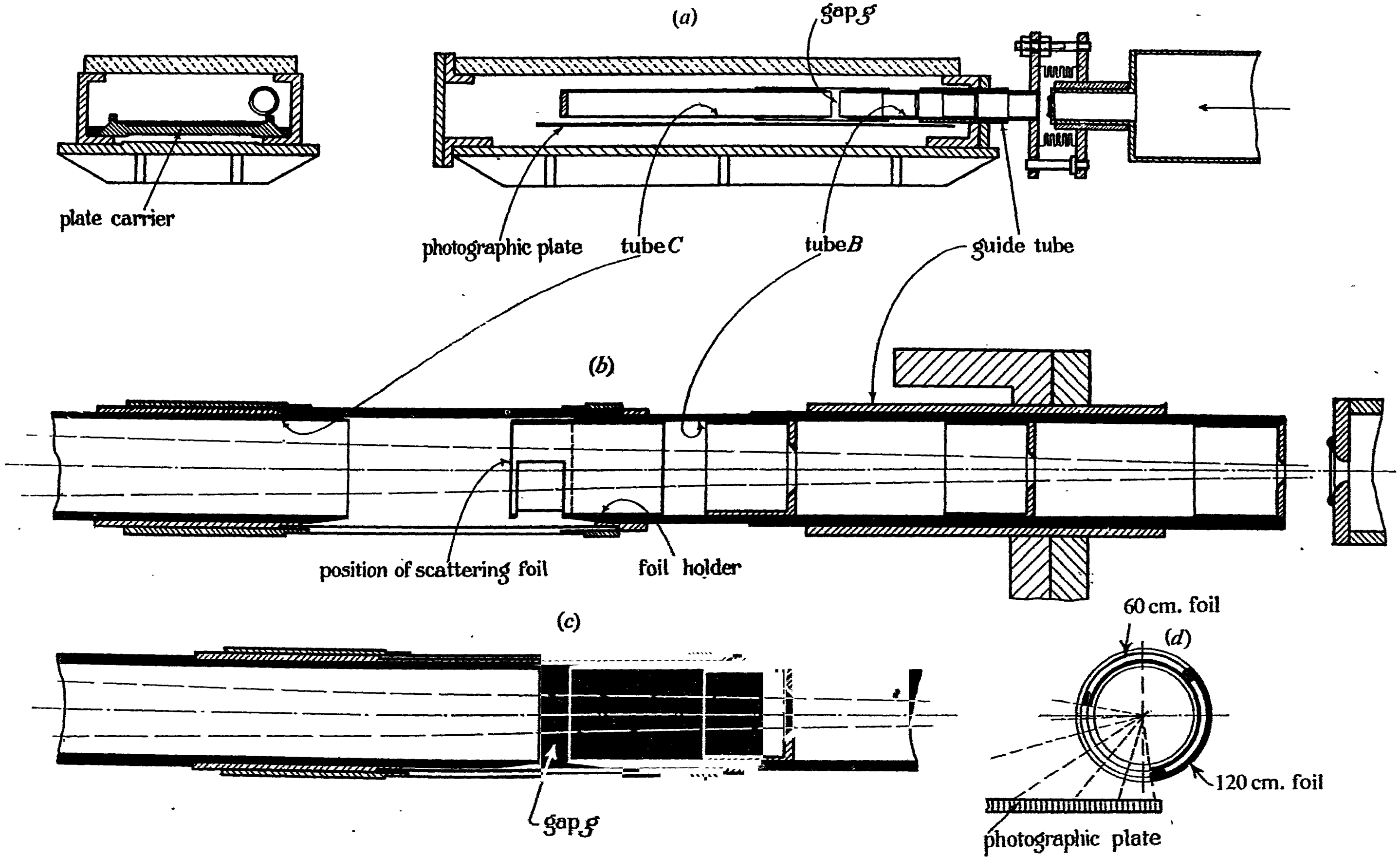


FIGURE 3. Detailed drawing of the scattering camera. (a) Side and end elevation of camera arranged for scattering by gases; (b) detail of scattering system for foils; (c) detail of scattering system for gases; (d) detail of arrangement of additional stopping foils.

This also avoids any bending of the photographic plate. For convenience in the microscopic examination the plates are made of glass only 1 mm. thick, and the tension in the gelatine may bend such a large thin plate if it is unconstrained. This is most likely to happen during the evacuation of the camera before the admission of the scattering gas, when the emulsion loses water by evaporation and tends to contract.

Before making an exposure, the loaded plate carrier is inserted into the camera, and the dark slide withdrawn through the rear end which is then covered with a brass plate, the joint being made gas-tight with a rubber gasket. The camera box is then exhausted and filled with the gas under investigation, through a side tube. Care is necessary to prevent the emulsion from drying out too rigorously, since this is apt to make the thick emulsions peel off the glass. At the same time it is important to have no water vapour in the scattering volume. It is found that a phosphorus pentoxide boat at the far end of tube C is adequate to keep the gas in the scattering volume dry, the aluminium foil preventing rapid diffusion of water vapour from the emulsion into the scattering volume. The camera is charged by successively filling with the gas and exhausting to the vapour pressure of water until the concentration of air has reached a sufficiently small value.

For experiments with target elements in the form of thin foils, the tubes B and C are withdrawn to increase the width of the gap between them to 38 mm. A foil holder, shown in figure 3*b*, is then inserted into the open end of tube B. The holder carries the scattering foil stretched across its end, perpendicular to the axis of the beam. The holder is cut away to allow the scattered particles to be observed in the backward as well as the forward directions. With this arrangement exposures with thin targets of aluminium and boron have been made, and from them we can determine the scattering in the angular range from 30 to 80° and from 100 to 145°. This restriction in the angular range is a serious disadvantage, and in future work the design of the foil holder will be modified to allow the scattering over the whole angular range to be examined.

3. GEOMETRY OF THE APPARATUS

(*a*) *Determination of intensity of scattered beam*

The essential features of the geometry of the apparatus are shown in figure 4 in which h represents the height of the axis of the beam above the plane of the surface of the emulsion. Suppose that around a point in the plate corresponding to an angle of scattering θ, a small area, A, contains $N_{K,\theta}$ tracks of a particular type K. Let the distance along the path of the particles from the centre of the scattering volume to the centre of the area A be r. Then the solid angle subtended by A at the centre of the scattering volume is $\frac{A}{4\pi r^2}\frac{h}{r}$. The scattering volume is given by la, where a is the area cross-section of the beam, and l is the length of the beam which can be seen from A through the gap g. If s denotes the length of the gap, $l = \frac{\rho'}{\rho}s$, where ρ' and ρ

are the distances of A, in the plane defined by A and the axis of the beam, from the axis of the beam and from the inner wall of the tube B respectively. The factor ρ_1/ρ clearly depends only on the Y co-ordinate of A (see figure 4). It is nearly equal to unity and varies only slowly with Y. To a first approximation the nuclear cross-section per unit solid angle at the angle θ, $\sigma_{K,\theta}$, for the process leading to particles of type K, may be written

$$\sigma_{K,\theta} = \frac{4\pi}{N_0 nsah} \frac{r^3 N_{K,\theta}}{A}, \tag{1}$$

where N_0 is the number of primary particles which have passed down the tube B in the beam during the exposure, and n is the number of nuclei per c.c. of the gas in the camera. Thus the quantity $r^3 N_{K,\theta}/A$, when corrected by the factor ρ'/ρ for the variation of the scattering volume with the value of Y at the point at which the count of $N_{K,\theta}$ is determined, will give the variation of the intensity of the scattered beam with angle. In this analysis the variation of the angle of scattering of the particles reaching a given point of the plate from different points in the scattering volume, and other corrections arising from the width of the gap g, and from the finite cross-section and divergence of the beam have been neglected. We have also assumed that the two tubes B and C are in perfect alinement with one another, and that the edges of the tubes at the gap g are infinitely thin. The experimental tests described below show that these assumptions give a sufficiently close approximation to the actual conditions. A more refined analysis is given in Appendix 1.

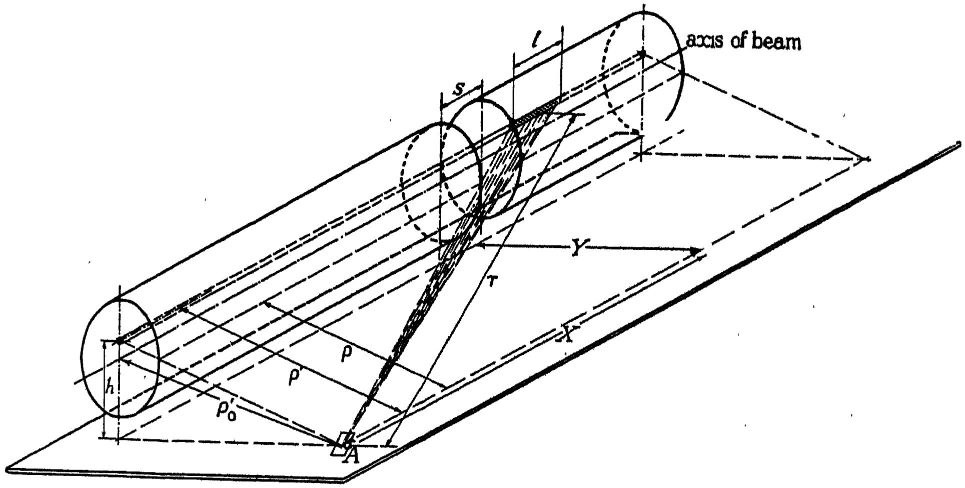


FIGURE 4. Geometry of the scattering system.

The analysis shows that the density of tracks at points in the plate corresponding to a given angle of scattering, θ, must be expected to vary as $1/r^3$. This leads to a very rapid variation of density across the plate. Thus at small values of r the plates are visibly blackened. With the exposures normally employed this blackening extends to a distance of about 2 cm. from the gap at small angles of scattering, and the form of this smudge gives a striking picture of the variation of the intensity of scattering with angle. Counting can be carried out at values of r ranging from 4 to 20 cm., corresponding to a ratio of 125 to 1 in density. This variation gives a considerable latitude in the value of the exposure required to produce satisfactory plates and allows the scattering at large and small angles to be measured on the same plate in spite of the rapid variation in the intensity with angle. The rapid variation of the density with r makes it desirable to determine r with an accuracy of the order of 0·1 mm. Thus with r equal to 5 cm., an error of 0·1 mm. in the determination of r gives rise to an error of 1·2 % in the value of the intensity. This is one-half of the statistical error arising from the number of tracks which it is usual to count at any particular value of the angle of scattering. It is therefore necessary to set the plate in the microscope stage, and to determine the position of the part of the plate being examined with this accuracy.

(*b*) *Determination of the angle of scattering*

Primary particles passing down the axis of the tube B and scattered through an angle θ so that they pass through the centre of gap g, trace out a cone which has its apex on the axis at a distance $R \cot \theta$ in front of the mid-plane of the gap g, where R is the internal radius of the tubes B and C. This cone cuts the plane of the emulsion in an hyperbola. The hyperbolae for values of θ at intervals of $2\frac{1}{2}°$ have been drawn on a large-scale map of the plate together with lines of constant distance from the centre of the scattering volume corresponding to any angle of scattering, a point which is taken to be the apex of the cone as described above. A simplified map of this type is shown in figure 5, and it will be seen that the lines of constant distance are not circles, since the position of the scattering volume varies with the angle of scattering.

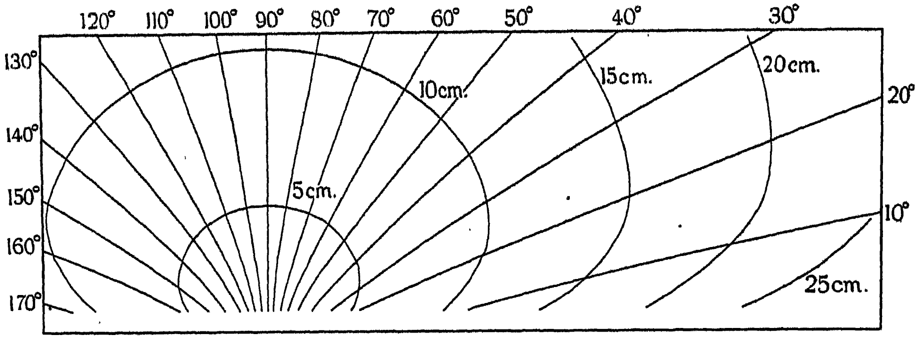


FIGURE 5. Map of plate showing lines of constant angle of scattering and of constant distance from the mean point of scattering. To avoid confusion the rectangular grid of dots is omitted.

In the exposures with foils the mean point of scattering is independent of the angle of scattering, so that lines of constant distance from this point are true circles on the photographic plate, and the lines of constant angle are hyperbolae corresponding to the intersection with the plane of the emulsion of cones with a common apex in the foil. A separate map is constructed for the plates exposed with the scattering foils appropriate to these geometrical conditions.

4. EXAMINATION OF THE PLATES

(a) *Microscope equipment*

The method of processing the plates and measuring the tracks has been discussed in previous papers. Ilford half-tone plates were employed with an emulsion thickness of 50 or 100μ according to the range of the scattered particles. After development the plates were cut into twenty-four equal pieces, each 1¼ by 1⅝ in., and covered with a no. 0 cover-slip of thickness 0·1 mm., mounted on Canada balsam. They were examined with a binocular microscope with a 3 mm. apochromat immersion objective in conjunction with a × 6 eyepiece. Occasional use was made of × 12 eyepieces for special purposes.

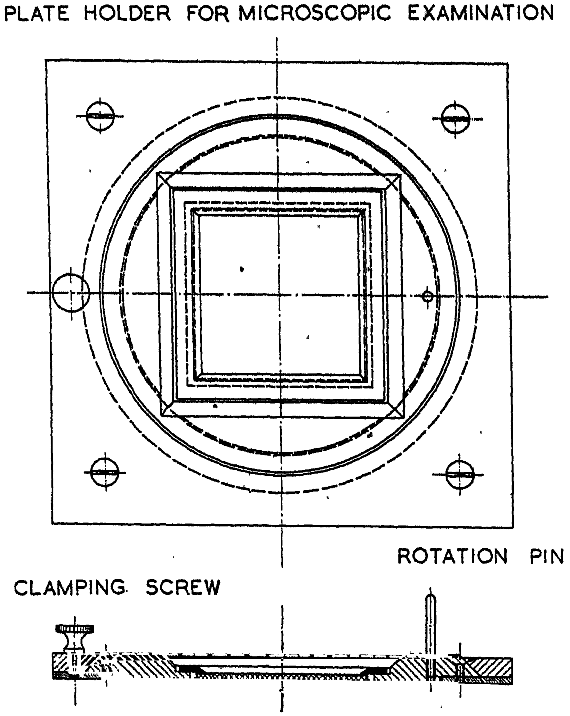


FIGURE 6. Plate holder for the microscopic examination of the plates.

In examining the particles corresponding to any value of the angle of scattering, the appropriate section of the plate is fitted in a small turntable, shown in section and plan in figure 6, which can be clamped to the mechanical stage of the microscope.

The turntable is rotated so that part of the hyperbola traced on the plate by particles scattered through the angle θ will be parallel to the X motion of the mechanical stage. The fiducial spot on the plate which lies nearest to the hyperbola for the angle of scattering θ is brought into the field of view of the microscope, and the perpendicular distance from this spot to the hyperbola is read off from the map (see figure 5). The Y motion of the mechanical stage is then shifted through this distance so that the tracks of particles scattered through the angle θ lie in the field of view. If the X wheel of the stage is now turned the angle of scattering of the tracks under observation is not altered.

(*b*) *Range distribution of the scattered particles*

To determine the distribution in range of the scattered particles, a part of the plate is chosen where the density of tracks is convenient, of the order of three or four per field of view, and the lengths of the tracks are measured against the eyepiece scale. The plate is scanned in the X direction, and the lengths of all tracks which originate in a defined central region of the field of view are measured. The speed of measurement varies greatly with the complexity of the range distribution of the scattered particles, but in simple cases it is of the order of one hundred an hour. The measurements are easier than in the case of the recoil protons in the investigation of neutron beams, since the tracks are all parallel and they start in the surface of the emulsion.

To convert the range of the scattered particles into cm. of standard air, the stopping power of the emulsion as determined in previous experiments is used; 1μ in the emulsion is taken to be equivalent to 1·46 mm. air. It is possible that small uncontrolled differences in the processes of manufacture may produce small changes in the stopping power for different batches of plates of the same type. We have no evidence for the existence of any significant differences of this kind and they are certainly small. Very accurate measurements of the energy of the particles have not appeared to be necessary in the experiments so far undertaken, and we have not thought it necessary to determine the stopping power of every batch of plates. The observed range must be corrected for the air equivalent thickness of the aluminium foil covering the gap and of the gas between the point of scattering and the point of observation on the plate.

(*c*) *Intensity measurements*

To measure the intensity of the scattered particles at a particular angle of scattering, in cases where the scattered particles are homogeneous in energy, the density of the tracks is determined by setting the plate as above, and turning the eyepiece scale until it is perpendicular to the tracks. The mechanical stage is then moved slowly in the X direction through, say, 4 mm., whilst the tracks crossing the scale are counted. The X motion is then made to return to its original setting and the Y motion is moved through a distance corresponding to the width of the eyepiece scale (0·1 mm.), and the plate is again scanned in the X direction through 4 mm. These operations are repeated until the number of tracks counted at this particular angle is sufficiently great to reduce the statistical error to the desired value. With the exposures normally used the variation in the angle of scattering over an area containing 1000 tracks is negligible. The value of r, the distance from the centre of the scattering volume to the centre of the scanned area, is read off from the map with an accuracy of 0·1 mm. Local small irregularities in the surface of the gelatine may cause errors in the determination of the intensity by changing the angle of incidence of the tracks on the surface of the emulsion. To avoid these errors at least four areas are scanned in the way described, at different values of r.

The above methods are satisfactory when the density of tracks is of the order of 200 per mm.^{2}. For higher densities a × 12 eyepiece is used instead of the usual × 6, and only those tracks are counted which cross the central region of the eyepiece scale. For the highest densities the × 12 eyepiece is used in conjunction with a graticule giving a rectangular grid in the field of view such as is used for blood counts. With the plate stationary, a map is made of the field of view from which the number of tracks starting in a definite area can be counted. The plate is then moved and another area is mapped in the same way. This method is very laborious, but the usual method is uncertain at high densities because the tracks sometimes cross one another and are difficult to disentangle, and it is sometimes useful to be able to deal with areas of high density.

When the scattered particles at a particular angle are not homogeneous, the above methods may still be employed if the tracks belonging to the different groups can be distinguished by inspection without measuring the length of each one. If, however, the range distribution is complex it becomes necessary to measure the length of every track starting within a definite area of the plate. Figure 7, for example, shows the distribution in energy of the particles from oxygen and nitrogen bombarded with 7 MeV deuterons. The two lower-range groups in the first case are sufficiently resolved, and can be counted without the measurement of the individual tracks, whereas in the second case the length of every track must be determined.

The nature of the particles producing the different peaks is given in the caption to figure 7, the identification of the different groups as protons, deuterons and α-particles having been confirmed by grain counts in a way to be described in a later paper. The results show the satisfactory resolving power achieved which appears to be considerably superior to that obtained with most of the best counter methods employed hitherto with primary beams of fast deuterons (see, for example, Humphreys & Watson 1941), and equal to that given by the expansion chamber (Schultz & Watson 1940).

At large values of r, the angles of incidence of the scattered particles on the surface of the emulsion become smaller. As a result there are a number of tracks which suffer small-angle deflexions towards the surface and leave the emulsion before coming to the end of their range. Such tracks can usually be distinguished in that they have both ends in the surface of the emulsion. If the particles are homogeneous in energy this effect leads to an apparent increase in the straggling of the range distribution curve, but does not affect the determination of the intensity of the scattered beam. The effect is most serious with slow particles and rapidly diminishes in importance as the energy increases, so that with the faster particles smaller angles of approach can be tolerated.

At angles of scattering less than ~ 10° the finite thickness of the chamfered edge of the tubes *B* and *C* leads to a decrease in the effective scattering volume which depends on the precise form of the edges of the tubes, and the observed density of tracks becomes correspondingly unreliable. With the apparatus in its present form, measurements cannot be made, therefore, below 10° and above 170°.

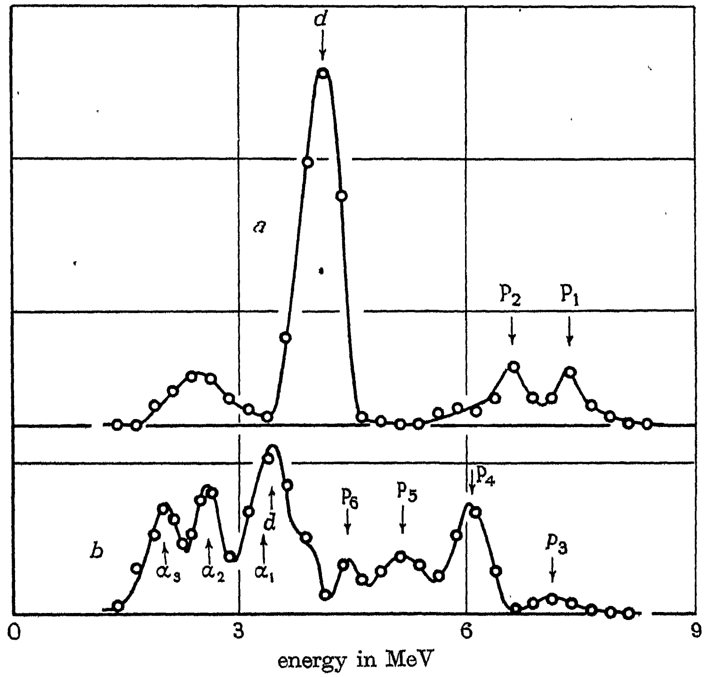


FIGURE 7. Particles from the bombardment of the light elements by 7 MeV deuterons: (*a*) oxygen at 45°, (*b*) nitrogen at 105°. The observed distribution in range has been transformed to a distribution in energy on the assumption that all the particles are protons. The peaks marked '*d*' are due to elastically scattered deuterons. The peaks P_1 and P_2 are due to protons from the reaction O^{16} (d, p) O^{17}, the O^{17} being left in the ground and first excited states respectively. P_3, P_4, P_5 and P_6 are protons from the reaction N^{14} (d, p) N^{15}, the N^{15} being left in excited states. The peaks α_1, α_2 and α_3 are α-particles from the reaction N^{14} (d, α) C^{12}, the C^{12} being left in the ground or excited states.

B. TESTS OF THE METHOD

5. GENERAL

A purely qualitative examination of the plates shows that some of the experimental difficulties have been overcome. In working with protons we have found no recoil tracks in the emulsion due to neutrons generated by the large ion currents in the tank, nor do the associated γ-rays and low-energy neutrons produce any appreciable increase in the general background of grains against which the tracks have to be distinguished. In work with deuterons the plate was shielded from neutrons coming directly from the tank of the cyclotron by boxes of wax placed round the scattering camera. Nevertheless, we find recoil tracks in appreciable numbers, due, we believe, to a localized source of neutrons produced by the impact of the deuteron beam on the brass in the neighbourhood of the stop s_2 (see figure 2). In future experiments we propose to cover the brass in this region with a foil of a heavy element. The tracks due to particles scattered by the gas or foil must all start in the surface of the emulsion and travel in a known direction. Very few of the neutron recoil tracks fulfil these conditions so that they do not lead to any confusion.

The number of tracks in 'wrong directions' due to particles scattered by the gas and then by the walls of the camera box is extremely small and gives no difficulty.

6. ANGULAR VARIATION OF THE ENERGY OF ELASTICALLY SCATTERED PROTONS

A satisfactory check on certain aspects of the method is provided by an investigation of the variation with the angle of scattering of the energy of protons, scattered elastically by the light elements. The energy, E, of particles of initial energy E_0 and mass m, after being scattered elastically through the angle θ by a nucleus of mass M, is given by the equation:

$$(M+m)E - 2m\sqrt{(EE_0)}\cos\theta - (M-m)E_0 = 0. \tag{2}$$

This relation can best be tested by means of the plates showing the scattering of hydrogen, deuterium and helium. The hydrogen exposure was taken with commercial hydrogen from a cylinder, the camera being filled to a pressure of 1 atm. The scattered protons have sufficient energy to penetrate the aluminium foil over the gap at angles of scattering less than 60° and give measurable tracks below 50°. The density of tracks on this plate is rather high, so that scattering at small angles has to be investigated at the greatest possible values of r. This, taken in conjunction with the high gas pressure and the fact that the gas contained a small quantity of oxygen, causes an unusually large apparent straggling in the observed range distribution at small values of θ. Above 25°, however, the proportion of tracks due to scattering from oxygen falls to less than 20 %, and the measurements between 25 and 50° are satisfactory. The results are shown in figure 8 and are in good agreement with the relation (2).

The deuterium used for the first experiments with this gas was very impure, containing both light hydrogen and oxygen in appreciable quantities. As a consequence we find in general three resolved groups of tracks at any angle of scattering, θ, of which the range varies with angle of scattering in the manner shown in figure 9. The protons scattered from oxygen vary only slowly with the angle of scattering, and the line through the experimental points for this group (curve a) follows closely that predicted by equation (2) if M is put equal to 16. Curve b is drawn from the results of putting M equal to 2, in equation (2), and the appropriate experimental points for protons scattered by deuterium lie on this line within the limits of experimental error. Finally, curve c passes through points corresponding to a third group of particles. Curve c is calculated on the assumption that these tracks are due to deuterons thrown forward by the impact of the primary protons, and the agreement between the observed and calculated values shows that the assumption is correct.

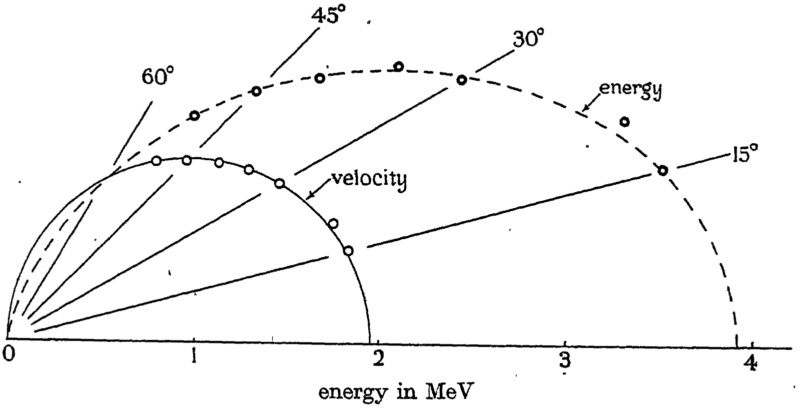


FIGURE 8. The energy and velocity of protons scattered by hydrogen at various angles as deduced from the observed range of the scattered particles. Circles give experimental points —the lines are calculated from the known primary proton energy.

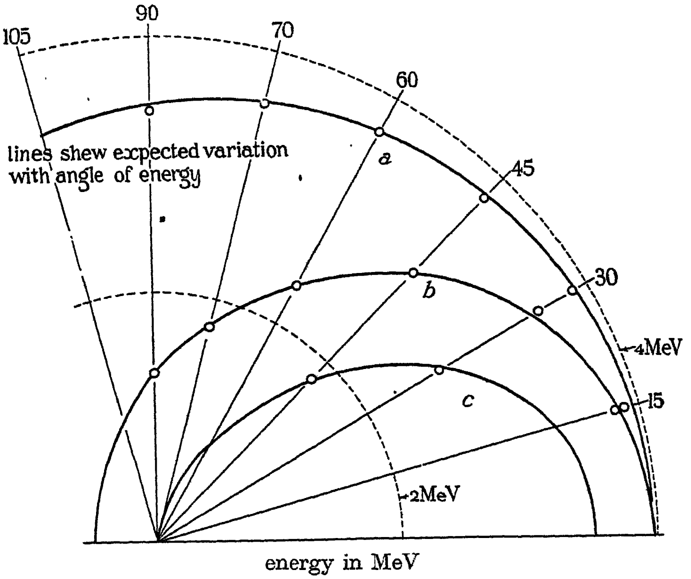


FIGURE 9. The variation with angle of the energy of particles from the impact of 4 MeV protons on deuterium containing oxygen as an impurity; (a) protons elastically scattered from oxygen, (b) protons elastically scattered from deuterium, (c) projected deuterons from the impact of protons on deuterium. Circles are experimental points. The lines show the expected variation calculated from the known primary proton energy.

Similar experiments with helium as the scattering gas also give close agreement between the experimentally observed values and those predicted by equation (2) when M is put equal to 4.

These measurements demonstrate that the observed tracks arise from single scattering in the gas and that the effects, on the observed range distributions, of multiple scattering and of scattering from the edges of stops, are negligible. They give a very direct check on the range-energy relation for the emulsion, and illustrate the power of the method by the ease with which the processes responsible for the different scattered groups can be identified.

7. VARIATION OF TRACK DENSITY WITH DISTANCE FROM THE POINT OF SCATTERING

It follows from equation (1) that the density of tracks for a constant value of θ should vary as $1/r^3$ when allowance has been made for the slight variation of the scattering volume with the Y co-ordinate of the point of observation. This relation has been verified in a large number of cases. In the case of the scattering of 4 MeV protons by hydrogen, experiments were made over a particular large range of r, from $r = 7\cdot6$ to $r = 18\cdot1$ cm. These measurements were made by two observers using the methods already discussed. The results are shown in figure 10, in which the

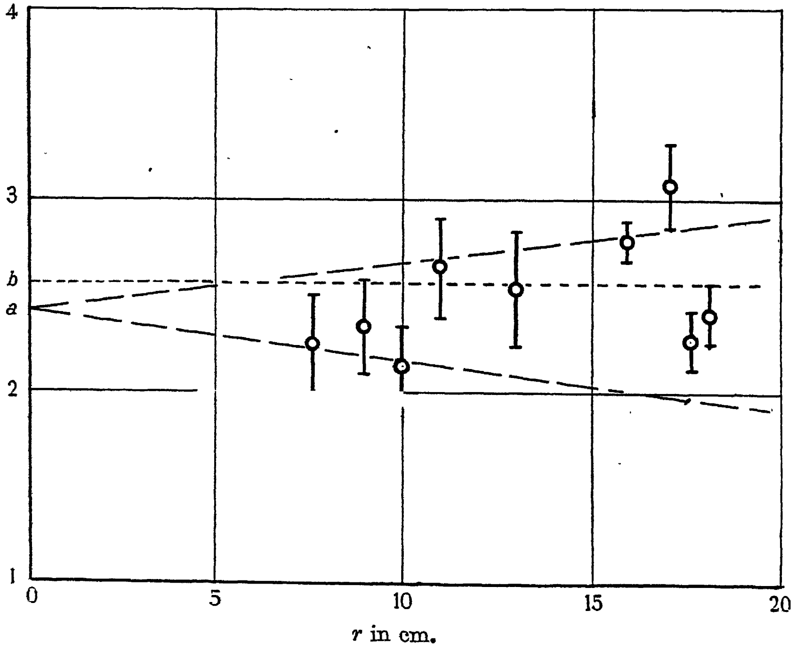


FIGURE 10. Proton-proton scattering. Diagram to illustrate the errors introduced into the determination of the scattered intensity due to uneven surface of the emulsion.

corrected value of the quantity r^3n/A is plotted against r. It will be seen that the differences from the mean are rather larger than the expected statistical fluctuations but do not show systematic change with r.

8. ACCURACY OF THE INTENSITY DETERMINATIONS

In the measurements represented in figure 10, and in some other cases, we have found discrepancies between the results at a given value of the angle of scattering, but at different values of r, which are larger than would be expected from statistical fluctuations. We ascribe this to irregularities in the surface of the emulsion. If, owing to these irregularities, the surface at any point is inclined at an angle δ then the particles impinge at a glancing angle $(h/r)+\delta$ on the surface. The intensity deduced from the counts, on the assumption that this angle is h/r, will be in error by a fraction $\delta r/h$. With $\delta = \pm\frac{1}{2}°$ this leads to errors of 20 % at 20 cm. and 10 % at 10 cm.

This is about the extreme discrepancy experienced. The error introduced is proportional to r and is shown in figure 10 graphically. To reduce this source of error, measurements of intensity are always taken at several different values of r, but the probable error of the mean result is still considerably greater than it would be from statistical fluctuation alone.

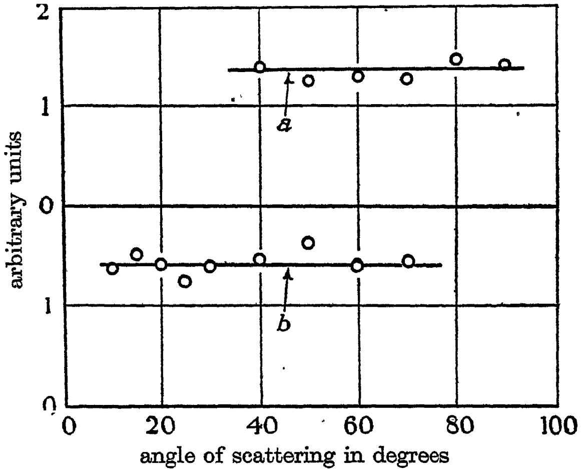


FIGURE 11. Elastic scattering of 4 MeV protons from (*a*) iodine and (*b*) argon, expressed at any angle as a ratio of the observed intensity to that expected from the Rutherford relationship.

Where measurements have been made over a wide range of r and where this is the main source of error, the observations should be given a weight proportional to $1/r^2$ in computing the mean. This leads to a slightly different mean from that deduced by weighting the observations according to the number of tracks counted. These two means are shown in figure 10. In most of our observations the range of r employed is not so great, and a simple mean of the results at different distances may be used.

A very satisfactory test of the accuracy of the determination of the intensity of the scattered beam and of the general performance of the apparatus is provided by experiments on the scattering of protons and deuterons by the heavy elements, where we must expect the intensity of the scattered beam to follow the Rutherford relationship, $I \propto \operatorname{cosec}^4 \theta/2$.

Figure 11 shows the results of the experiments on the scattering of 4·0 MeV protons by iodine in the form of CH_3I vapour, and by argon. The observed intensity of the scattering at any angle is plotted as a ratio to that expected if the variation in the intensity followed the relation $I = K \operatorname{cosec}^4 \theta/2$ precisely. The results were obtained with an earlier form of the camera in which there was no provision for observations at angles of scattering greater than 100°. The scattering is Coulombian in the range examined within the limits of statistical error. The reliability of the method is thus confirmed.

9. ANGULAR RESOLVING POWER OF THE APPARATUS AND ACCURACY OF DETERMINATION OF THE ANGLES OF SCATTERING

The angular resolving power of the apparatus as calculated from the geometry of the arrangement varies with both r and θ, and is given to a first approximation by the equation

$$\Delta\theta = \frac{S \sin\theta}{r} \text{ radians.} \tag{3}$$

This gives $\Delta\theta = 20'$ at $r = 20$ cm., $\theta = 15°$ and $\Delta\theta = 3{\cdot}4°$ at $r = 5$ cm., $\theta = 90°$. It is satisfactory that the resolving power is greater when θ is small and the variation of the intensity with θ most rapid. As is shown in Appendix 1 (*c*) these values of the resolving power are not attained because of the divergence of the primary beam and of the small-angle secondary scattering of the particles by the gas filling the camera, as they move from the gap to the emulsion. A very simple and direct demonstration of the resolving power attained in practice is provided by experiments on the scattering of deuterons by protons.

It is easy to show that when a deuteron beam is scattered by hydrogen, the deuterons cannot be scattered through angles greater than 30°. We therefore expect to find a group of particles which disappears at this angle. The details of the intensity relations will be discussed in a later paper, and it will be shown that for an ideally sharp angular resolving power we should expect a variation of intensity of this group with angle of the type shown in curve *c*, figure 12. In fact we observe a variation according to curve *b*. This smearing out of the sharp features of the angular distribution corresponds to a finite angular resolving power of the order of ± 2°, about half of which is due to small-angle scattering in the gas. This is amply sufficient for the resolution of detail occurring in the angular distribution of scattered particles from other elements, although it may lead to errors in intensity measurements when the intensity is varying rapidly with angle as in Rutherford scattering at small angles. This is discussed in Appendix 2 (*e*).

10. GENERAL REMARKS

As a result of our experience in the course of the present work, we have made certain minor improvements in the apparatus. Experiments with deuterons, to be described subsequently, showed the importance of making experiments at a number of different values of the energy of the primary beam using any particular scattering gas. We have therefore made provision for the introduction of stopping foils into the primary beam so that the primary energy can be immediately changed to any one of a number of values.

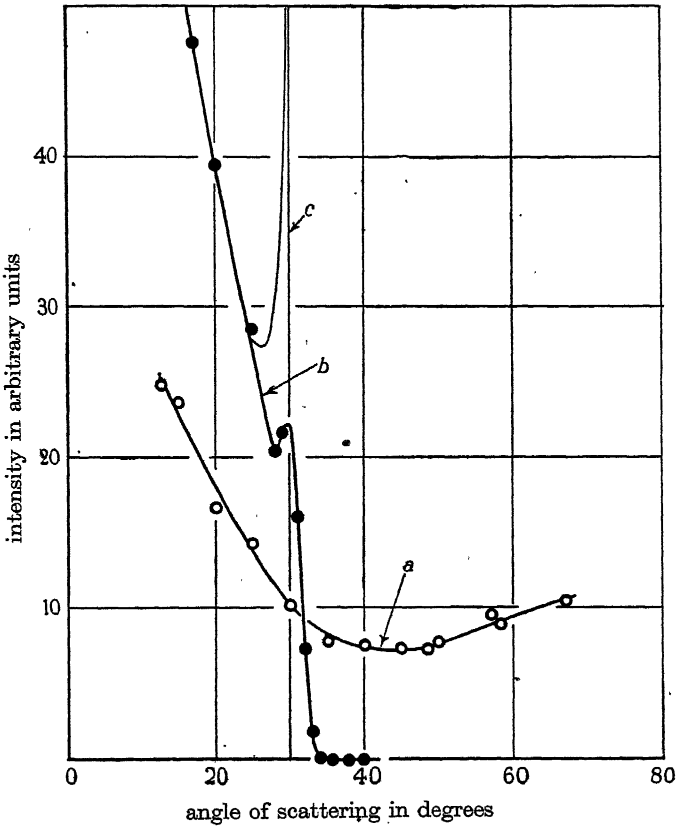


FIGURE 12. Distribution in angle of the intensity of particles from the impact of 7 MeV deuterons on hydrogen. (*a*) Projected protons. (*b*) Scattered deuterons as observed. (*c*) Expected distribution of scattered deuterons as calculated from the observed proton distribution assuming infinite resolving power. The discrepancy between the observed and calculated deuteron distributions is due to the failure to observe some of the deuterons of which the residual range in the emulsion is small and to the finite resolving power of the apparatus.

In the apparatus as described, the scattering gas fills the whole camera and is not recovered after the exposure. For experiments with rare gases, such as separated isotopes, provision must be made for the recovery of the gas, and in general a sufficient quantity to fill the apparatus to the required pressure is not available. It is therefore desirable to replace the aluminium foil hitherto used to cover the gap, g, by a mica window so that the collision space between the tubes B and C can be made gas-tight. In order to avoid pressure differences between the two sides of the mica, it is designed to fill the camera with a common gas such as hydrogen at the same pressure as the rare gas filling the collision space.

The new cyclotrons now in course of construction will yield deuteron beams with an energy of the order of from 50 to 100 MeV. One must expect new and unforeseen problems and difficulties in passing from experiments with 10 MeV to those with 100 MeV particles, but it seems reasonable to assume that the methods described may still be applicable, without major modifications, at considerably higher energies than those of which we have experience.

Two of us (A. N. M. and C. F. P.) have pleasure in acknowledging our indebtedness to Professor A. M. Tyndall for placing at our disposal the facilities of the H. H. Wills Physics Laboratory, University of Bristol, where the construction of the camera and the measurements on the plates were carried out.

APPENDIX 1

A MORE ACCURATE THEORY OF THE SCATTERING CAMERA

(a) *Effect of the finite cross-section of the beam*

The beam is assumed to be accurately parallel to the common axis of the tubes B and C. An element of the beam at a height h above the plate contributes a number of tracks at the point of observation A which is proportional to $F = \frac{h}{r^3}\rho'/\rho$ in the notation on p. 8. Introducing co-ordinates X and Y in the surface of the plate as shown in figure 4 we have,

$$\left.\begin{aligned} r &= \rho'/\rho\sqrt{(X^2+\rho^2)} = \rho'\sqrt{(1+X^2/\rho^2)}, \\ \text{and}\qquad F &= \frac{h}{(1+X^2/\rho^2)^{\frac{3}{2}}}\frac{1}{\rho\rho'^2}. \end{aligned}\right\} \tag{1}$$

Now we introduce co-ordinates τ, ϕ, fixing the position of an element of the beam relative to the axis of the scattering tube (figure 13). Then using the suffix 0 to designate the value of any quantity for an element on the axis, it is easily shown that to the second order in τ,

$$\left.\begin{aligned} h/h_0 = 1+\tau\sin(\phi-\alpha), \quad & \rho'/\rho'_0 = 1-\frac{\tau\cos\phi}{\rho'_0}+\frac{\tau^2\sin^2\phi}{2\rho'_0}, \\ \rho/\rho_0 = 1+\frac{\tau^2\sin^2\phi}{2R\rho_0}, \quad & 1+\frac{X^2}{\rho^2} = \left(1+\frac{X^2}{\rho_0^2}\right)\left(1-\frac{\tau^2\sin^2\phi\cos^2\theta}{R\rho'_0}\right), \end{aligned}\right\} \tag{2}$$

where $\alpha = \tan^{-1} h_0/Y_0$, R is the internal radius of the scattering tube, and θ is the angle of scattering. Inserting these values in (1) and developing as far as the terms in τ^2 we obtain

$$F/F_0 = 1+\frac{\tau}{h_0}\sin(\phi-\alpha)+\frac{2\tau\cos\phi}{\rho_0'}+\frac{2\tau^2}{h_0\rho_0'}\sin\overline{\phi-\alpha}\cos\phi +\frac{\tau^2}{\rho_0'^2}(4\cos^2\phi-1)+\frac{\tau^2}{2R\rho_0'}(3\cos^2\theta-1)\sin^2\phi,$$

where θ is the angle of scattering.

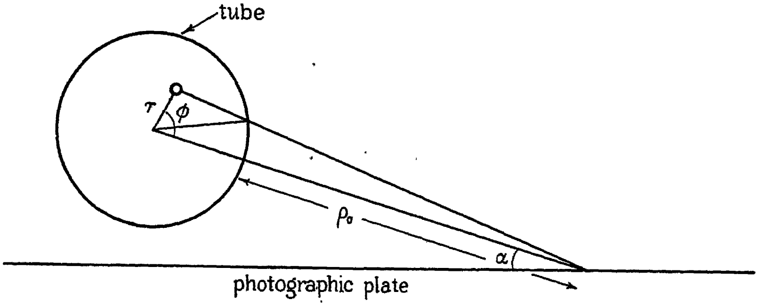


FIGURE 13. Cross-section of scattering tube and photographic plate.

If we average this over the cross-section of the beam then the first order terms depend on the distance of the centre of gravity of the beam from the axis. The first term is independent of the co-ordinate X, Y of the point of observation on the plate and therefore does not affect relative intensity determinations. The second depends on ρ_0' which varies from 5 to 10 cm., so that the extreme error in relative intensity measurements will be $2\xi/100$, where ξ is the displacement of the centre of gravity of the beam from the axis in mm. This reaches 2 % if $\xi = 1$ mm.

To investigate the effect of the second order of terms we assume that the beam is accurately centred on the axis, is circular in cross-section, and of uniform intensity across the section. Then we average first with respect to ϕ and then τ.

If we average with respect to ϕ, then the first order terms in τ disappear and the second order terms become

$$-\frac{\tau^2}{h_0\rho_0'}\sin\alpha+\frac{\tau^2}{\rho_0'^2}+\frac{\tau^2}{R\rho_0'}\frac{3\cos^2\theta-1}{4}.$$

Now average τ between 0 and T, the radius of the beam, and we get

$$\bar{F}/F_0 = 1-\frac{T^2}{2h_0\rho_0'}\sin\alpha+\frac{T^2}{2\rho_0'^2}+\frac{T^2}{R\rho_0'}\frac{3\cos^2\theta-1}{8}. \qquad (3)$$

Here $F_0 = \frac{h_0\rho_0'}{r_0^3\rho_0}$, and therefore includes the corrections for the variation of the scattering volume with Y.

The correction terms in (3) are small. Thus taking typical values, $T = 0{\cdot}2$, $R = 0{\cdot}75$, $h_0 = 1{\cdot}45$, $\rho_0, \rho_0' = 10$, all in cm., the correction terms would not exceed 0·2 % and may therefore be safely neglected.

For large values of θ the scattering occurs at a considerable distance from the defining stop so that the beam may be appreciably broadened. Suppose we take $T = 0{\cdot}5$ cm., then the correction terms amount to 1·2 % at $r = 10$ cm., and 2·4 % at $r = 5$ cm. This is small enough to be neglected.

(b) *Effect of the finite width of the gap g*

The above calculation was made for an infinitely narrow gap. For a fixed point of observation the different parts of a gap of finite width will give different values of X in equation (1) and we have to replace the value of F for the centre of the gap by the mean value over the width, s. This will change the observed intensity by a factor

$$\left(1+\frac{X^2}{\rho^2}\right)^{\frac{3}{2}} \frac{1}{s} \int_{-s/2}^{+s/2} \frac{dx}{\left[1+\left(\frac{X+x}{\rho}\right)^2\right]^{\frac{3}{2}}} = 1-\frac{s^2}{8}\frac{\rho^2-4X^2}{(\rho^2+X^2)^2}$$

$$\approx 1+\frac{s^2}{8r^2}(4\cos^2\theta-1) \approx 1+10^{-3} \quad \text{at } r = 10 \text{ cm., } s = 0{\cdot}5 \text{ cm.}$$

(c) *Angular resolving power*

There are several causes of slight variations of the angle of scattering at a fixed point of observation on the plate. First the angular divergence of the primary beam is 1° as given by the geometry of the stop system. It will be rather more than this on account of small-angle scattering in the mica window s and in the gas, but this effect is small. Secondly, the width of the gap g gives a spread in angle of the scattered beam given by

$$\Delta\theta = \frac{s}{r}\sin\theta. \tag{4}$$

Besides this, however, there is a variation in θ which is due to the finite diameter of the primary beam, and would be present even with an infinitely thin gap. We have $\rho/X = \tan\theta$. Now ρ varies across the beam according to equation (2). If η is the extreme value of $r\sin\phi$ this leads to an extreme variation of θ from the value on the axis given by

$$\Delta\theta = \frac{1}{2}\frac{\eta^2\cos^2\theta}{XR} = \frac{\eta^2}{2rR}\cos\theta.$$

This will be less important than (4) at all angles greater than $\theta_c \sim \eta^2/2sR \sim 2°$. This effect can therefore be neglected.

Finally, we must consider small-angle scattering in the gas between the scattering volume and the point of observation. The root mean square angle of scattering is given in degrees by $Z\sqrt{(PL)}/2E$, where Z is the atomic number of the gas, P the pressure in atmospheres, L the length traversed in cm., and E the energy of the particle in MeV. This leads to values of the order of 1°.

APPENDIX 2

(*a*) *Variations in the energy of the primary beam corresponding to different angles of scattering*

In experiments with gases the mean point of scattering corresponding to any particular angle of scattering, θ, varies with θ. Owing to the stopping power of the gas filling the camera, it follows that there is a slow decrease in the energy of the primary beam at the mean point of scattering as θ increases. Even in unfavourable cases, however, this variation is small, and it can be made negligible, when desired, by reducing the pressure of the gas in the camera and increasing the exposure time.

To take a particular example, in one exposure with acetylene at a pressure of 30 cm. of mercury in the camera, the energy of the deuteron beam on the vacuum side of the mica windows (see figure 3) was 7·4. The energy at the mean point of scattering varied from 6·7 MeV for $\theta = 15°$ to 6·4 MeV for $\theta = 165°$, a range of 0·3 MeV. For this experiment the exposure time was 8 min. with 5 μA in the main cyclotron beam reaching the collecting plate (see figure 1). In similar conditions with 10 cm. of gas in the camera, the exposure time necessary to give the same density of tracks would have been 24 min., and the variation in the primary energy from 7·06 MeV for $\theta = 15°$ to 6·96 MeV for $\theta = 165°$, a range of 0·1 MeV. In work with deuterons an even greater reduction in the pressure could have been made without leading to an appreciable number of neutron recoils in the emulsion, and in work with protons in this range of energy neutron disturbance is negligible. The necessary exposure times could of course be reduced by an increase in the main currents in the cyclotron.

(*b*) *Penetration of the edges of the gap*

This would give rise to a tail on the short-range side of any peak in the range distribution. The edges of the gap are of course not quite sharp, the tubes B and C being feathered externally at an angle of 5° to a thickness of about 0·01 cm. This corresponds to about 40 cm. of air, or 27 microscope divisions. This is sufficient to stop all scattered particles and most of the disintegration protons. Nevertheless, at angles of observation other than 90° a certain amount of penetration will occur in the way shown in figure 14. Particles which enter the brass at a distance x cm. from the edge will go through a thickness $x/\cos\theta$ cm. of brass, a thickness equivalent to $2700x/\cos\theta$ microscope scale divisions. Thus if N particles go through the gap of width 5 mm. and form the main peak of the distribution, the particles which pass through the edge of the gap will give a flat tail with a height of $4N\cos\theta/2700$ tracks per microscope division. Thus a peak containing 700 tracks would have a tail with one track per microscope division. The observed tail is more intense than this owing to straggling in the emulsion and mismeasurement of a few tracks.

Particles with an initial range greater than 40 cm. of air will be able to penetrate the scattering tubes over a greater area. For them a similar calculation gives a tail of height $4N \sin \theta/270$ tracks per microscope division. This will stretch from a range 27 microscope divisions less than that of the main peak down to zero range.

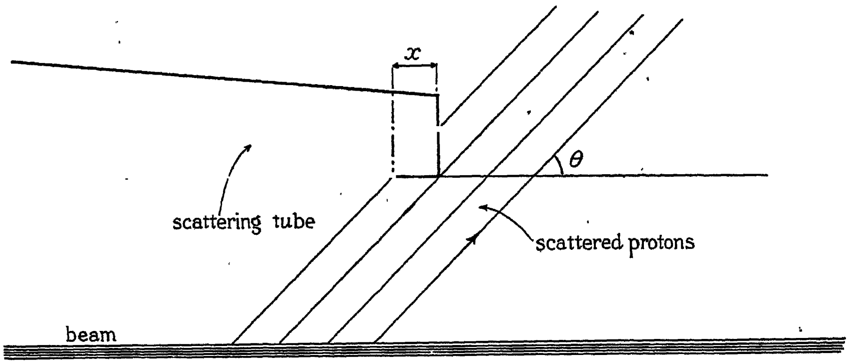


FIGURE 14. Effect of penetration of the edges of the slit system.

(c) *Inhomogeneity and fluctuation in the energy of the primary beam*

This would show up in our experiments as a broadening of the elastically scattered peak. We have not observed any such effects in experiments with 4 MeV protons and conclude that the primary beam is homogeneous and constant in energy within the resolving power of the method in this case. With 7 MeV deuterons there is evidence for a spread in energy of the primary beam of about 0·5 MeV.

(d) *Effect of stray magnetic field at the scattering camera*

For convenience in working the scattering camera was placed in a horizontal plane. The vertical component of the stray magnetic field from the cyclotron will therefore cause a curvature of the tracks in this plane. Measurements of the stray field were made with a search coil and these show that in the region of the camera the stray magnetic field is about $\frac{1}{20}$ of the main field. Since the latter gives a radius curvature of 40 cm. in the tank, the radius of curvature of elastically scattered particles in the camera will be 8 m. Disintegration particles, or inelastically scattered particles, will have a greater or less curvature corresponding to their values of $H\rho$.

Now the length of path of the scattered particles in the camera is of the order of 10–20 cm. and in this distance the above curvature will lead to the particle turning through an angle of about 1°. The error in the assumed angle of scattering will be $\frac{1}{2}$ of this, or 0·5°, which is less than the error in angle arising from other causes.

(e) *Errors in the measurement of intensity at small angles*

(i) *Errors arising from bad alinement.* If the axis of the beam does not coincide with the axis of the scattering tubes, or if from any other cause there is an error $\Delta\theta$ in the determination of the angle of scattering θ, the rapid variation of intensity with angle at small angles will give rise to serious errors in the apparent angular distribution. Thus for Rutherford scattering $I = \text{const.} \operatorname{cosec}^4 \frac{1}{2}\theta$, so that

$$\Delta I/I = 2\cot\tfrac{1}{2}\theta\,\Delta\theta = 4\Delta\theta/\theta$$

at small angles. This would amount to a 40 % error at 20°, for an error of 2° in the measurement of θ. The best test of this type of error is the satisfactory agreement found with Rutherford scattering.

(ii) *Errors from finite resolving power.* A similar but smaller effect will occur because the tracks at any point of observation have been scattered through a finite range of angles from $\theta-\Delta\theta$ to $\theta+\Delta\theta$. Again assuming Rutherford scattering this will give rise to an error of the order $4(\Delta\theta/\theta)^2$ at small angles. For $\Delta\theta = 2°$ this amounts to 2·5 % error at 20°.

(iii) *Finite thickness of the feathered edge, and bad alinement of tubes B and C.* If t is the thickness of the edges of the tubes at the gap, then the effective length of the gap is reduced by an amount $t\cot\theta$ for particles passing through the axis of the tube and rather more for others. This introduces an error $t\cot\theta/s$.

If the axes of the tubes B and C are parallel but displaced by a distance t' in the horizontal plane, then the above formula will hold with t changed to $t \pm t'$, the sign to be taken changing at $\theta = 90°$. If t, t' are assumed to be about 0·1 mm. this gives of the order of 5 % error at 20°.

REFERENCES

Bower, J. C. & Burcham, W. E. 1939 *Proc. Roy. Soc.* A, **173**, 379.
Humphreys, R. F. & Watson, W. W. 1941 *Phys. Rev.* **60**, 542.
Kuerti, G. & Wilkins, T. R. 1940 *Phys. Rev.* **57**, 1081.
Pollard, E. & Watson, W. W. 1940 *Phys. Rev.* **58**, 12.
Powell, C. F. 1940 *Nature, Lond.*, **145**, 155.
Powell, C. F. 1943 *Proc. Roy. Soc.* A, **181**, 344.
Powell, C. F. & Fertel, G. E. F. 1939 *Nature, Lond.*, **144**, 115.
Powell, C. F., May, A. N., Chadwick, F. & Pickavance, T. G. 1940 *Nature, Lond.*, **145**, 893.
Schultz, H. L., Davidson, W. L. & Ott, L. H. 1940 *Phys. Rev.* **58**, 1043.
Schultz, H. L. & Watson, W. W. 1940 *Phys. Rev.* **58**, 1047.
Wilkins, T. R. 1941 *Phys. Rev.* **60**, 365.
Wilkins, T. R. & Kuerti, G. 1939 *Phys. Rev.* **55**, 1134.
Wilkins, T. R. & Kuerti, G. 1940 *Phys. Rev.* **57**, 1083.

Long duration of the Balmer spectrum in excited hydrogen

By Lord Rayleigh, F.R.S.

(*Received* 19 *November* 1943)

[Plate 1]

The light of a condenser discharge through hydrogen, emitting the Balmer series of lines, persists for much longer than current theories and experimental determinations of the time of relaxation of the hydrogen atom would lead one to anticipate. These determinations indicate a time of the order of 10^{-8} sec. for the intensity to diminish in the ratio $e:1$. It is shown in the present paper that in fact, if hydrogen made luminous by a powerful discharge is blown out of the electric field by its own expansion, the time in question is under some conditions a thousand times more. The discrepancy is not cleared up, but the aim has been to present the evidence as clearly as possible for criticism.

This paper deals with the question of how long the luminosity of hydrogen (Balmer spectrum) lasts after the electric discharge has ceased to pass through the gas.

Results obtained with metallic vapours (Strutt 1914*a*, *b*) suggest that in their case the duration is much longer than most theoretical discussions of the subject would lead one to anticipate, since mercury vapour, for example, can be made to pass in the luminous condition down long tubes, and it can be shown that the same applies to other metallic vapours.

Theoretical views as to what hydrogen ought to do are much more definite than in the case of other elements because of the comparative simplicity of the case from the standpoint of wave mechanics. In the present paper the behaviour of hydrogen is examined by experiment. Earlier experiments had led to the doubt whether the usual view could be sustained (Strutt 1917), and the present more definite investigation has confirmed the earlier one. The experiments appear to indicate definitely enough that in this work the fall in intensity in the ratio $e:1$ occupies not 10^{-8} sec., as found in other investigations, but 10^{-5} sec. So enormous a factor (1000 times) shows clearly that the discrepancy is fundamental, and cannot be met by any criticism on points of detail.

A summary by S. Dushman of the literature on this subject will be found in Taylor & Glasstone (1941). The first and in some respects the most direct experiments on this subject (Wien 1919, 1921) were made by taking advantage of the luminosity of the canal rays, and their velocity measured by magnetic and electrostatic deviation. They gave the values $2{\cdot}3 \times 10^{-8}$ sec. for H_β and H_γ. Slack (1926) has determined the duration of the $2P$ state in hydrogen, using excitation with a voltage of about 10·2 which could only excite the first line of the Lyman series. Without going into details of his method, it may be said that the radiation was observed photoelectrically, and that its duration was brought into comparison with the time of oscillation of a high-frequency circuit of 10^7 cyc./sec. He concluded that this intensity fell in the ratio $e:1$ in $1\cdot2\times10^{-8}$ sec.

Condon & Shortley (1935, p. 135) have given a quantum-mechanical calculation of the duration of the various components of the H_α line. This line is a complex of seven components. These authors give the mean lives (the reciprocals of the total transition probabilities) of each level of the three initial configurations as:

configuration	mean life
3*s*	16×10^{-8} sec.
3*p*	$0\cdot54\times10^{-8}$ sec.
3*d*	$1\cdot56\times10^{-8}$ sec.

I have not attempted to follow in detail the complex train of reasoning by which this result is arrived at. It is in general agreement with the time (10^{-8} sec.) calculated by the classical (Hertzian) method (Lorentz 1909) and seems to be strongly supported by the general success of the quantum-mechanical treatment of spectra. This quantum theory points clearly to periods not much more than 10^{-8} sec., and such experimental evidence as is available points to about the same result.

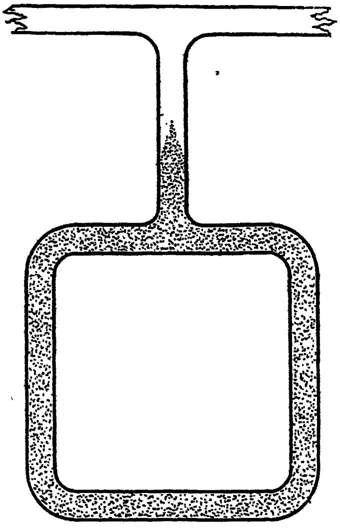


FIGURE 1

The general plan of the present experiments is as follows: The electrodeless discharge is produced by induction in a tube having the form of a square ring. This tube had an annexe or side tube coming out at right angles to one side and of the same bore, through which it was exhausted. The tube is vertical in figure 1.

Each discharge is of very short duration and of high energy, capable of heating the small mass of rarified gas to a high temperature with corresponding rise of pressure. The gas squirts out into the vertical exit above mentioned, out of the path of the current, remaining luminous (figure 2, plate 1). It there shows the three hydrogen lines H_α, H_β and H_γ. By observing the vertical exit tube with a revolving mirror the upward progress of the luminosity can be traced and its velocity determined and the decay of luminosity at various distances along this tube can be measured independently. These data make it possible to determine the time in which the initial luminosity of the gas decays in any given ratio.

In detail the arrangements are these. The discharge without electrodes is induced in an endless tube about 13 cm. square and 1·5 cm. bore (figure 1). Inside this is the primary (exciting) coil consisting of 12 turns or less of thick copper wire, each 10 cm. square. This coil is earthed in the middle. Two condensers and a spark gap are in series with it, the spark gap being connected to a high-potential transformer. The sparks are at the rate of 100 per sec., which is the doubled frequency of the alternating supply. The spark length between balls was 8·5 mm., corresponding to say 28 kV. The aggregate capacity of the two condensers in series was $7{\cdot}25 \times 10^{-3}\,\mu$F. This makes the energy of charge $2{\cdot}94 \times 10^7$ ergs; thus the output of the condensers is 294 W. The power dissipated in the ring discharge was measured by immersing the discharge tube in an oil calorimeter. It was found to be about 100 W, say 0·3 of the output of the condensers. The rest, no doubt, goes almost entirely into the spark in the condenser circuit. The 100 W represents 1 joule or 0·24 cal. at each discharge. The discharges are, however, of very short duration, with relatively long intervals between. The duration may be examined by means of a rotating mirror, applied either to the spark gap in the primary exciting current, or to the low-pressure discharge.

Each discharge, as will be seen, occupies only a very short space of time, and the momentary rise of temperature in the gas can be calculated from the measured energy of a discharge combined with a knowledge of the heat capacity. Thus, taking the volume of the tube as 79 c.c. and the hydrogen pressure as 0·2 mm., the mass of the gas is $1{\cdot}88 \times 10^{-6}$ g. The specific heat may be taken as 3·4 cal./g. and the thermal capacity will be $6{\cdot}4 \times 10^{-6}$ cal. Thus one discharge will cause a rise of $0{\cdot}24/6{\cdot}4 \times 10^{-6}$ degree or $3{\cdot}76 \times 10^4$ degrees. A great instantaneous increase of pressure is therefore to be expected and the expansion producing the luminous jet is amply accounted for.

It is possible that some who have had experience of the perverse way in which condenser discharges go to earth by any path except where the experimenter wishes, may be suspicious that the prolongation of the luminosity into side vessels, as in figure 2 (plate 1), is not a genuine effect of transport of a stream of gas having residual luminosity, but that it represents an electric current gone astray to earth. For many reasons I am satisfied that this is not the case. Such stray discharges often occur if a single condenser only is used. The stray discharge is in part superposed on the genuine effect of the jet of luminous gas. They are, however, easily distinguished by several criteria. Stray discharges show a complex spectrum, the so-called secondary (molecular) spectrum of hydrogen, whereas the luminous jets only show the Balmer lines on a black background, and are dominated by the fiery red of the H_α line. They are not more than 6 cm. in length. If the channel divides, as in figure 3, the fiery luminous jets divide also (figure 4, plate 1), but the stray discharge goes down on one side only to the air pump, which forms in effect an earthed electrode. It also extends through any length of tubing, in contrast to the luminous jets, which only go a short distance.

The best way of getting rid of these effects is to use Lodge's arrangement with two condensers in series with the spark gap, and the exciting coil in the other symmetrical position between the condensers. The exciting coil may be earthed at the middle, though this is scarcely necessary. The stray discharges then disappear, but the genuine effect remains.

If the light from the spark gap is examined, a long train of oscillations is found, with little damping, the period being (with one of the arrangements used) $1{\cdot}29 \times 10^{-6}$ sec. When the secondary discharge started, the long train of regular oscillations was reduced to 3 or 4 faintly discernible, by far the greater part of the light being in the first discharge or pilot spark.

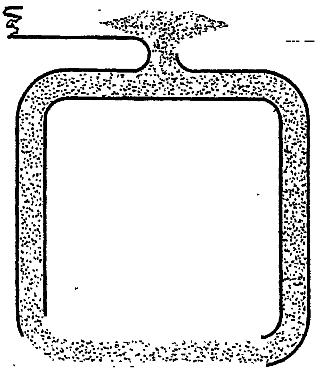


FIGURE 3

If the light of the induced discharge using nitrogen in the square discharge tube is examined, then $1{\cdot}93 \times 10^{-6}$ sec. of total duration, or about 1½ periods, is found. If hydrogen is used instead of nitrogen a complication ensues. In hydrogen the light is persistent owing to the gas remaining luminous, and does not go down to nothing at the minima of the oscillating current. No oscillations can in fact be detected, and if the mirror turns very fast, as is necessary to resolve the oscillations with nitrogen, the light in hydrogen is diluted to the vanishing point, and there is nothing to be seen or photographed. By using a rapid ($f/1{\cdot}9$) lens and a larger mirror a luminous trail can be photographed. The initial image with the mirror turning slowly (merely to separate the individual discharges) was 0·1 mm. wide. By turning the mirror at 29 rev./sec., the image was expanded and was traceable for a length of 0·5 mm., representing $1{\cdot}8 \times 10^{-5}$ sec. The images (enlarged about 23 times) are represented in figures 5 and 6 (plate 1).

This experiment, only carried out with difficulty, seems to give some indication of a discontinuity of intensity, the later decay being slower. However, it shows clearly enough that the duration is much more than can be reconciled with the received view that the time is of the order 10^{-8} sec. It is generally confirmatory of the more satisfactory experiment described below.

A point of interest is the very high instantaneous value of the power dissipated in the discharge tube. Taking the duration of each discharge of the condenser as $1{\cdot}93 \times 10^{-6}$ sec., and the number of discharges as 100 per sec., the discharges occupy only $1{\cdot}93 \times 10^{-4}$ of the whole time: and the power dissipated during the progress of an individual discharge is $100/1{\cdot}93 \times 10^{-4}$ W, or 520 kW.

This enormous figure for the power dissipated in a vacuum discharge tube excited in this way came as a surprise to me, and is perhaps not generally realized. It helps to account for the distinctive character of condenser discharges in spectroscopy. Since the duration of $1{\cdot}93 \times 10^{-6}$ sec. includes $1\frac{1}{2}$ complete oscillations, it is evident that the true maximum power must be several times that calculated.

In order to measure the velocity, a slit 0·5 mm. wide was placed in front of the tube; this was photographed with a Dalmeyer cinema lens of $f/1{\cdot}9$ aperture and Ilford Astra VII plates which are specially sensitive to H_α light. The image on the plate was 0·085 actual size. A revolving mirror (silver on glass) was placed in front of the lens, it being large enough to fill the lens aperture with light. The rate of rotation was 33 rev./sec. The displacement of the image by a given rotation was measured directly on the ground glass of the camera, giving the mirror a small rotation by means of a distance piece of known thickness. It was found that 1 cm. displacement of the image corresponded to 0·07 radian or $3{\cdot}39 \times 10^{-4}$ sec. of time.

The discharges occurring at the rate of 100 per sec., a considerable number of them were photographed at each exposure. The direction of the slit was recorded by rotating the mirror very slowly through one revolution by hand. Such a slow rotation serves to separate the discharges but does not give any indication of the velocity of the gas impelled along the tube by a discharge, which velocity is, in these conditions, to be regarded as infinite in comparison.

Having made a preliminary exposure with slow rotation, the motor is started and the mirror run at a standard speed of 33 rev./sec. Another exposure is given, superposed on the former one. On development, numerous images from both exposures are seen, those at slow speed being almost uniformly spaced, and the others interpolated at random among them (figure 7, enlarged, plate 1).

A specially clear photograph selected from another plate is reproduced in figure 8, (plate 1). It is easy to distinguish the exposures made with the mirror running at full speed, which are inclined at a small but quite perceptible angle to those made with the mirror going very slowly. This obliquity is the basis of measurement. Measurement is made to the sharp left-hand edge of the photographed image. This edge records the time at which the luminous gas first reaches a given height. The obliquity was measured as 4·5° by means of a low-power microscope, making settings on the lines by a spider thread in the rotating eyepiece. The result is regarded as correct within 0·5°. As the luminosity progresses along the tube the mirror rotates, and the image is displaced through an appreciable distance.

On the scale of the original photograph 1 cm. horizontal represents $3{\cdot}39 \times 10^{-4}$ sec. of time; 1 cm. vertical represents 11·8 cm. of distance. The observed slope of 4·5° is equivalent to 1 cm. vertical for 0·08 cm. horizontal.

This gives for the velocity with which the luminosity advances $11{\cdot}8/0{\cdot}08 \times 3{\cdot}39 \times 10^{-4}$ or $4{\cdot}36 \times 10^{5}$ cm./sec. (It may be mentioned for comparison that the velocity $4{\cdot}36 \times 10^{5}$ cm./sec. above calculated would be the velocity of atomic hydrogen in temperature equilibrium at 768° abs. or 495° C.)

The luminosity in the vertical exit tube (figures 1, 2) was examined as follows:

A plate was prepared with a line of small (2 mm.) holes at intervals of 1 cm. A neutral glass reducing the light nearly in the ratio $e:1$ was placed over the first of these holes, and this was placed over the entrance of the vertical exit tube (figures 1, 2). The second hole, with no opaque glass, was brighter, but the difference was less in each successive hole, and at the fifth hole (4 cm. from the first) the unscreened luminosity was judged equal to the initial luminosity reduced by the opaque glass.

It appears then that the luminosity is reduced in the ratio $e:1$ after a distance of 4 cm.

At the velocity $4{\cdot}36 \times 10^{5}$ cm./sec. this represents a time interval of $0{\cdot}92 \times 10^{-5}$ sec., or, say, about 10^{-5} sec. It is not asserted that after another 10^{-5} sec. the intensity will have again fallen in the same ratio. The indication is rather that the fall will be more rapid.

As will be seen by inspection of figure 2 on plate 1, the luminosity in the vertical side branch tapers out to a conical end. This seems to show that it is lost by diffusion of the luminous centres to the walls of the tube, when they become extinguished. This cause of extinction is superposed upon the decay of luminosity in the body of the gas. Owing to the complication introduced by this wall effect, it is scarcely feasible to study the exact law of decay by photometric measurements along the tube.

The rate of decay is not exactly the same for the various lines of the Balmer series. A greenish blue fringe is sometimes seen near the limit of the exuded luminosity, and examination with a direct-vision prism shows that this is due to the tongue of luminosity being a little longer in the light of H_β than in that of H_α or H_γ. This is somewhat strange, as one would expect regular progress along the series. The difference of behaviour of the lines is not very marked, and indeed has not always been noticeable. I have not determined what are the best conditions for observing it.

It is clear from the above that for the Balmer lines in general at least 10^{-5} sec. is required for the luminosity to sink to $1/e$ of its actual initial value, under the conditions used. This is a thousand times greater than the period (10^{-8} sec.) usually assigned from theoretical considerations, and from Wien's experiments on the luminosity in the track of positive rays.

While no attempt is made to account for this, the evidence has been reproduced in a way which, it is hoped, will give the reader the opportunity to form his own opinion of it, and to weigh it against the theoretical objections.

Experimental work which seems out of harmony with established fashions of thought does not easily find its way into books, and is apt to be soon forgotten. The writers of text-books and summaries of literature do not know how to deal with it, and find the easiest solution is not dealing with it at all. I hardly expect any other result in the present case, but it seems a duty to put the observed facts on record. It is believed that the photographic evidence which is here reproduced will put the reader in almost as good a position to judge of it as if he had carried it out himself.

REFERENCES

Condon & Shortley 1935 *Theory of Atomic Spectra.* Cambridge.
Lorentz, H. A. 1909 *Theory of Electrons*, p. 259.
Slack, F. G. 1926 *Phys. Rev.* **28**, 1–12.
Strutt, R. J. (Rayleigh) 1914*a* *Proc. Roy. Soc.* A, **90**, 364–372.
Strutt, R. J. (Rayleigh) 1914*b* *Proc. Roy. Soc.* A, **91**, 92–103.
Strutt, R. J. (Rayleigh) 1917 *Proc. Roy. Soc.* A, **94**, 88–99.
Taylor, H. S. & Glasstone, S. 1941 *Physical Chemistry*, **1**, chap. II. London.
Wien, W. 1919 *Ann. Phys., Lpz.*, **60**, 597.
Wien, W. 1921 *Ann. Phys., Lpz.*, **66**, 229.

EXPLANATION OF PLATE

FIGURE 2. Ring discharge induced in hydrogen at 0·2 mm. pressure. Note the extrusion of luminosity 6 cm. or so into the tube above.

FIGURE 4. Luminosity extruded into a T-tube instead of the straight vertical tube of figure 2. Note how it divides into two streams.

FIGURE 5. Image of a slit placed over the vertical exit tube. Hydrogen discharge.

FIGURE 6. Image of a slit in a revolving mirror. Note the drawing out of the image. Length of arrow denotes a duration of $1{\cdot}8 \times 10^{-5}$ sec. In nitrogen under the same conditions the image is not perceptibly drawn out.

FIGURE 7. Jet of extruded luminosity. The longer vertical images, of which there are five at equal intervals, are taken with the mirror revolving very slowly. The shorter images inclined at 4·5° to the former are taken with the mirror revolving at speed. The slope measures the speed with which the front of the extruded luminous jet moves forward as about 4×10^5 cm./sec.

FIGURE 8. A similar picture from another plate giving about the same result.

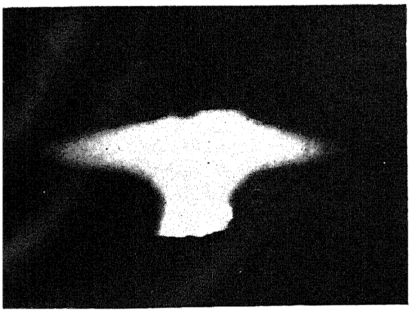


FIGURE 4

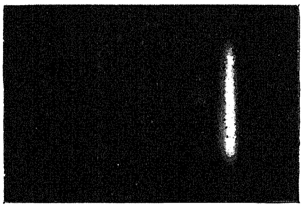


FIGURE 5

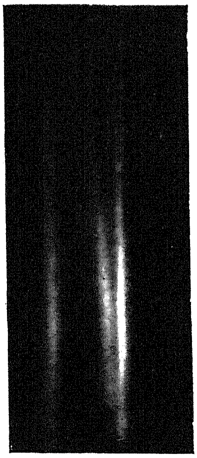


FIGURE 8

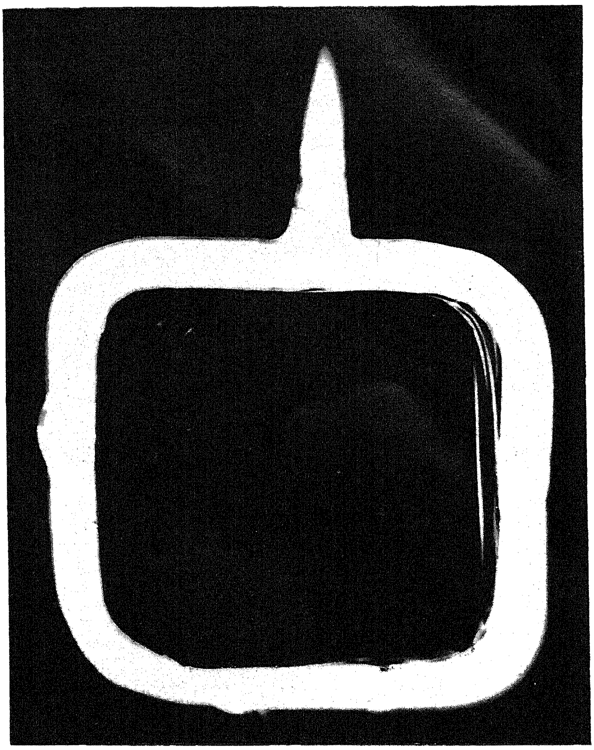


FIGURE 2

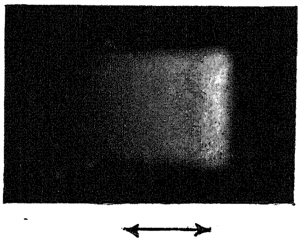


FIGURE 6

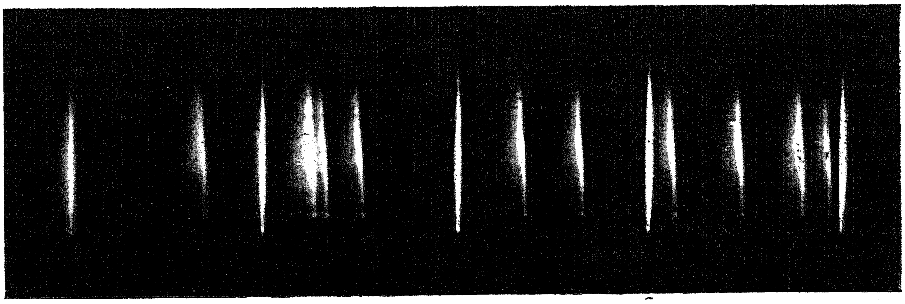


FIGURE 7

The thermal decomposition of acetone

By J. R. E. Smith, D.Phil., and C. N. Hinshelwood, F.R.S

(*Received* 2 *December* 1943)

Inhibition by propylene shows that free radical chains occur in the thermal decomposition of acetone.

Quantitative investigations permit of conclusions about the mechanism of the chain processes, which are compared and contrasted with those occurring in other decomposition reactions.

A free radical chain mechanism for the homogeneous thermal decomposition of acetone was postulated by Rice & Herzfeld (1934). This paper gives direct evidence that part of the reaction does in fact proceed by such a mechanism. However, in several respects the properties of the reaction, as hitherto known, differ from those of unambiguous chain reactions. The reaction rate does not vary with the surface-volume ratio (Hutchison & Hinshelwood 1926; Winkler & Hinshelwood 1935), nor when free radicals are added by decomposing mercury dimethyl (Rice, Radowskas & Lewis 1934). Again the analogous photolysis has a quantum yield of less than one. Finally, Staveley & Hinshelwood (1937) found that the decomposition is not sensitive to added nitric oxide, whose radical absorbing property had in other cases proved a definite indication of chain mechanisms. This evidence led some to conclude that the disruption of acetone liberates only molecules with saturated valencies. On the other hand, Rice & Polly (1938) reported an inhibition of the reaction by propylene, which would combine with free radicals; and more recently Rice found more, but less definite, evidence of a free radical chain mechanism from the analysis of the products at low pressures (private communication). Since useful data concerning the decomposition of other compounds have been given by a detailed study of the inhibitory effect of propylene (Smith & Hinshelwood 1942), a similar approach has now been made to the problem of the mechanism of the acetone decomposition.

The decomposition of acetone occurs at a measurable rate at temperatures at which propylene itself reacts with gradual increase in pressure. Consequently the pressure increase of acetone in the presence of propylene bears no constant ratio to the amount of acetone decomposed, and the reaction was followed in all cases by analyses for acetone using the iodoform method (see Winkler & Hinshelwood 1935).

The rate of the reaction was measured by the reciprocal of the time for 50 % decomposition, and was found to be considerably reduced by the addition of propylene (table 1, figure 1). This provides unequivocal evidence of the existence of a free radical chain mechanism. As with other decompositions the rate is reduced to a well defined limit with no evidence of a secondary catalysis even with large amounts of propylene.

TABLE 1. THE RATE OF DECOMPOSITION OF ACETONE IN THE PRESENCE OF PROPYLENE AT 570° C

partial pressure of propylene in mm.	$1/t_{\frac{1}{2}} \times 10^3$			
	50 mm. acetone	100 mm. acetone	200 mm. acetone	300 mm. acetone
0	1·64	1·90	2·08	1·98
2	1·20	—	—	—
5	0·96	1·35	1·60	—
10	0·81	1·14	1·42	1·67
20	0·70	0·96	1·32	1·46
50	0·60	0·72	1·05	1·21
100	0·53	0·69	0·91	1·03
150	—	—	0·80	—
170	—	0·58	—	0·83
200	—	—	0·76	—
estimated limiting rate	0·50	0·55	0·60	0·60

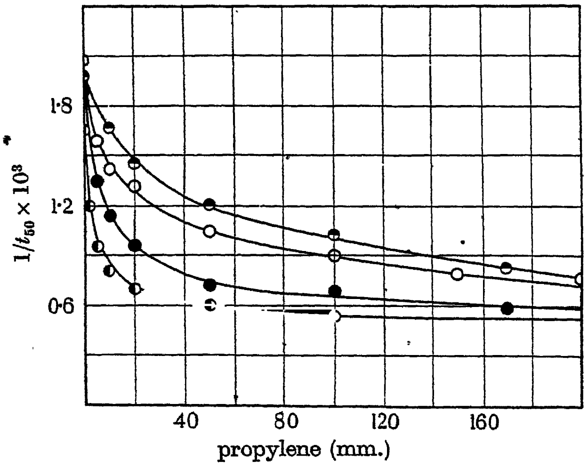


FIGURE 1. Inhibition of the acetone decomposition by propylene at 570° C. (Vertically divided circles, 50 mm. of acetone; black circles, 100 mm.; open circles, 200 mm.; horizontally divided circles, 300 mm.)

This appears to be in conflict with the apparent lack of inhibition of the reaction by nitric oxide (Staveley & Hinshelwood 1937). Consequently the effect of nitric oxide was reinvestigated (table 2, figure 2). An unmistakable inhibition is revealed. This is counterbalanced at large pressures of nitric oxide by a catalysis (as observed with acetaldehyde, Smith & Hinshelwood 1942). In the simpler decompositions of diethyl ether and propaldehyde it was definitely shown that the same limiting rate was produced by the two inhibitors, nitric oxide and propylene (Smith & Hinshelwood 1942). From figure 2 it is reasonable to estimate that with acetone, if it were not for the catalysis, the limiting rates for the two inhibitors would again coincide. Thus it is probable that both nitric oxide and propylene can largely inhibit the chain reaction, leaving in both cases the same residual reaction of a non-chain nature.

TABLE 2. THE RATE OF DECOMPOSITION OF 100 MM. OF ACETONE AT 570° C IN THE PRESENCE OF NITRIC OXIDE

partial pressure of nitric oxide in mm.	$1/t_{\frac{1}{2}} \times 10^3$
0·0	1·90
0·5	1·85
1·0	1·77
2·0	1·68
5	1·56
10	1·46
20	1·40
36	1·71

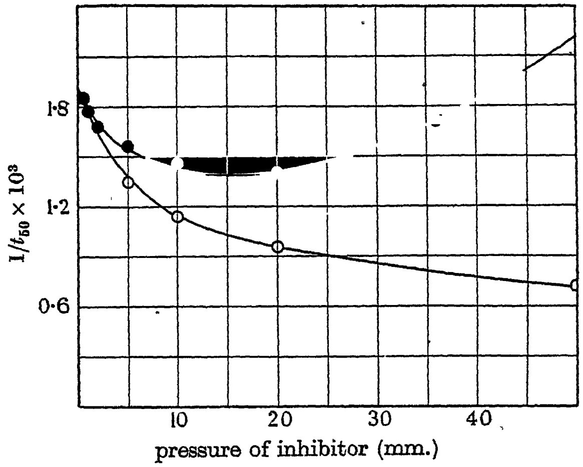


FIGURE 2. Inhibition of the acetone decomposition at 570° C by nitric oxide (black circles), and by propylene (open circles).

From table 1 and figure 1 it is apparent that the limiting value of the reaction rate for large amounts of propylene is nearly the same for different initial pressures of acetone. That is, the maximally inhibited reaction is of the first order. Further, the normal uninhibited reaction is also of the first order above 100 mm. pressure, as confirmed by table 1 (see also Winkler & Hinshelwood 1935). Consequently the difference between the rates of these two reactions, which represents the inhibitable portion of the reaction, is itself constant with varying initial pressure of acetone. Thus the acetone chain reaction is of the first order.

The amount of propylene required to suppress a given fraction of the inhibitable reaction increases as the initial pressure of acetone increases. This is shown by figure 1, or by table 3 in which is tabulated $(\rho-\rho_\infty)/(\rho_0-\rho_\infty)$ against the partial pressure of inhibitor for several initial pressures of acetone (ρ is the rate for a given amount of inhibitor, ρ_0 is the rate without inhibitor, and ρ_∞ is the limiting rate). The 'inhibition curves' are not superimposable.

A consideration of the steady state shows that the proposed chain reaction (given below) is consistent with the first order of the reaction if the chains are broken largely by step (4) rather than the recombination of two like radicals:

$$
\begin{aligned}
&(1)\quad CH_3COCH_3 &&\rightarrow 2CH_3 + CO,\\
&(2)\quad CH_3 + CH_3COCH_3 &&\rightarrow CH_4 + CH_2COCH_3,\\
&(3)\quad CH_2COCH_3 &&\rightarrow CH_3 + CH_2CO,\\
&(4)\quad CH_3 + CH_2COCH_3 &&\rightarrow C_2H_5COCH_3.
\end{aligned}
$$

That step (4) is the chain-breaking step would indicate a concentration of CH_2COCH_3 radicals comparable with that of CH_3 radicals. However, if the propylene attacks the CH_2COCH_3 radical, a steady state calculation shows that the 'inhibition curves' would coincide for different initial pressures of acetone. That the 'inhibition curves' are widely separated (see table 3) leads to the assumption that the propylene attacks the CH_3 radical. Apparently propylene combines more rapidly with the CH_3 radical than with more bulky radicals in comparable concentrations. This conclusion was also reached from a similar argument in the case of the decomposition of diethyl ether (Smith & Hinshelwood 1942).

Table 3. 'Inhibition-curve' data for acetone and propylene at 570° C

pressure of propylene in mm.	$(p-p_\infty)/(p_0-p_\infty)$ 50 mm. acetone	100 mm. acetone	200 mm. acetone	300 mm. acetone
0	1·00	1·00	1·00	1·00
2	0·61	—	—	—
5	0·40	0·60	0·68	—
10	0·27	0·44	0·55	0·78
20	0·18	0·30	0·48	0·62
50	0·09	0·13	0·30	0·44
100	0·03	0·10	0·21	0·31

In several respects the acetone reaction differs from other chain reactions. The reaction rate is not sensitive to change in the surface-volume ratio, nor to the free radicals from decomposing mercury dimethyl. Radiation cannot induce a quantum yield greater than one. Higher temperatures are required to produce comparable rates of reaction. Finally, very much more nitric oxide is required to produce appreciable inhibition than, for example, with the decomposition of diethyl ether. These facts would be explained if the mean chain length in the acetone chain decomposition were less than with most of the other decompositions studied. Then in acetone relatively more chain-breaking processes would occur while a given number of acetone molecules decomposed. Consequently alternative, competing chain-breaking processes due to inhibitor or surface would be less important, as would also be the addition of free radicals from an external source at a fixed rate. Moreover, higher temperatures would be necessary to supply a concentration of free radicals to maintain the reaction rate.

The activation energy of the reaction in the presence of large amounts of propylene is 62,600 cal. per mol. (table 4). Although this is 6000 cal. less than the activation energy of the normal uninhibited reaction (Winkler & Hinshelwood 1935), the reaction rate is still much greater than would be produced by a simple collisional activation in two squared terms. (For example, at 100 mm. pressure and 600° C the rate of maximally inhibited reaction is $2{\cdot}1 \times 10^{15}$ mol. cm.$^{-3}$ sec.$^{-1}$, whereas $Ze^{-E/RT}$ is only $7{\cdot}8 \times 10^{11}$.) Indeed, this is to be expected since the maximally inhibited reaction is of the first order. Thus the results for the maximally inhibited reaction are still consistent with the hypothesis of an internal rearrangement process in which many degrees of freedom contribute to the activation.

TABLE 4. THE TEMPERATURE COEFFICIENT OF THE RATE OF DECOMPOSITION OF 100 MM. OF ACETONE IN THE PRESENCE OF 100 MM. OF PROPYLENE

temp. °C	$t_{\frac{1}{2}}$ sec.
615	193
600	372
585	686
570	1306
555	2638

TABLE 5. RELATIVE EFFICIENCIES AS INHIBITORS OF PROPYLENE AND NITRIC OXIDE IN DIFFERENT DECOMPOSITIONS

reactant	ratio of the amounts of C_3H_6 and NO to produce a given inhibition	probable radicals absorbed
acetone	approx. 1	CH_3
acetaldehyde	approx. 5	CH_3
diethyl ether	14	CH_3, $CH_2OC_2H_5$
propaldehyde	100	C_2H_5

A comparison of the results for acetone with those for other decompositions shows that the relative efficiencies as inhibitors of nitric oxide and propylene depend very greatly on the particular decomposition (table 5). It was concluded from kinetic results with the decomposition of diethyl ether that propylene, in contrast with nitric oxide, combines more readily with the smaller CH_3 radical than with the larger $CH_2OC_2H_5$ radical. Again from table 5 the advantage of nitric oxide over propylene seems to be much smaller in combining with the CH_3 radical than with the somewhat larger C_2H_5 radical. It is perhaps surprising that the effect is as large as observed, but this seems the only explanation.

REFERENCES

Hobbs 1938 *Proc. Roy. Soc.* A, **167**, 456.
Hutchison & Hinshelwood 1926 *Proc. Roy. Soc.* A, **111**, 245.
Rice & Herzfeld 1934 *J. Amer. Chem. Soc.* **56**, 284.
Rice & Polly 1938 *J. Chem. Phys.* **6**, 273.
Rice, Radowskas & Lewis 1934 *J. Amer. Chem. Soc.* **56**, 2497.
Smith & Hinshelwood 1942 *Proc. Roy. Soc.* A, **180**, 237.
Staveley & Hinshelwood 1937 *J. Chem. Soc.* p. 1568.
Winkler & Hinshelwood 1935 *Proc. Roy. Soc.* A, **149**, 340.

On the flow of gas through a nozzle with velocities exceeding the speed of sound

By L. Fox, D.Phil. and R. V. Southwell, F.R.S.

(*Received* 16 *December* 1943)

An iterative method outlined by Green & Southwell (1944), but not applied in their paper for the reason that it makes no use of 'relaxation methods', is here applied to determine the supersonic regime for gas flowing irrotationally through a convergent-divergent nozzle, i.e. that (unique) regime in which the pressure and density of the gas decrease continuously in its passage from end to end.

Osborne Reynolds's approximate treatment of the problem (1886) assumed the velocity to be distributed uniformly over each cross-section, and in consequence found a unique value for the limiting mass-flow, whether the velocity be subsonic or supersonic downstream of the 'throat' (upstream it is always subsonic). Here, a more exact treatment shows that the supersonic value is very slightly (0·045 %) greater than the subsonic value, which Reynolds's theory overestimates by 0·083 %. The two regimes of course imply two different values of the pressure at exit, and for intermediate pressures (on the assumptions of this paper) there is no solution of the problem.

Even in the subsonic regime (when the mass flow is critical) velocities exceeding the local speeds of sound are attained in regions adjoining the nozzle walls near the throat.

INTRODUCTION

1. Osborne Reynolds (1886) showed that the discharge of gas through a convergent-divergent nozzle cannot exceed a certain limiting value, dependent on the starting conditions, however far the back-pressure is reduced. His treatment disregarded the effects of curvature of the stream-lines, assuming the velocity to be uniform in any section. On that understanding the limiting mass flow is attained when in the smallest section (or 'throat') the gas has a velocity equal to the local speed of sound.

In these circumstances (as was shown by Reynolds) either of two regimes can occur in sections downstream of the throat: either (*a*) after attaining its maximum expansion at the throat the gas may again contract (with increase of pressure) as it passes to the exit section, or (*b*) it may continue to expand, with a further decrease of pressure. No regime is possible of a type intermediate between (*a*) and (*b*), and accordingly it may be conjectured (by analogy with the theory of equilibrium of mechanical systems) that one or other of these two regimes is unstable.

2. When an attempt is made to improve the theory by a treatment which makes allowance for curvature of stream-lines, even in the case of laminar (two-dimensional) flow the problem is very difficult. Taylor & Sharman (1928) avoided its purely mathematical difficulties by having recourse to an electrical analogue—the flow of current in a conducting layer of varying thickness. By this means they could determine flow patterns so long as these entailed no velocity as great as the local speed of sound; but at that limit the method failed and its results became divergent—thus giving further indication of some form of instability.

Similar experience was reported in a recent paper (Green & Southwell 1944) which attacked the same (two-dimensional) problem by 'relaxation methods'. A technique devised for other problems ('liquidation' of residual errors by a use of 'relaxation patterns' appropriate to a regular net) proved successful as applied to flows involving no velocity above the local speed of sound; but it too broke down as the mass flow approached its limiting value (§ 1), in a manner again suggestive of some latent instability.

3. Now a regime that is unstable may yet be discoverable by calculation, as is shown (e.g.) in the theory of compression members (struts) and of 'free' streamlines; but it is not to be expected that it will be revealed by tentative methods such as are described in § 2, *because disturbances are implied in any inexact solution.* For this reason Green & Southwell, in §§ 11–12 and 22–24 of their paper, outlined an alternative (non-relaxational) method, of iterative type, which should serve to determine the supersonic as well as subsonic regimes. They did not proceed to results, as inappropriate to a paper concerned with relaxation methods.

This paper uses their alternative method to compute the supersonic flow pattern for a closely similar† shape of nozzle, also a limiting subsonic pattern in extension of their results. The work was done in part before, but for the most part after, Dr Green's departure for Australia. With one or two exceptions (which are explained) it confirms his results, and it completes his partial (i.e. subsonic) solution; but it leads to two additional and important conclusions, viz.

(1) that *two* limiting values of the mass flow in fact replace the single value given by Reynolds's less exact analysis: one is the upper limit under conditions which are subsonic both upstream and downstream of the 'throat'; the other is that value for which supersonic velocities occur at all downstream sections. The second value is about 0·045 % higher than the first, which Reynolds's theory overestimates by some 0·083 %.

(2) that at the lower value, while the velocity is subsonic both upstream and downstream, it is supersonic in two localized regions adjoining the walls at the throat.

The paper is confined strictly to computational aspects of the problem, being intended simply as a supplement to Green & Southwell (1944). Discussion of the physical bearing of the work is reserved for a subsequent paper, and on that account no reference is made to previous work other than that of Taylor & Sharman (1928).

I. OUTLINE OF THE ITERATIVE ('STRIP INTEGRATION') METHOD

4. In the earlier paper (Green & Southwell 1944) the gas was assumed to start from rest in a reservoir where its pressure is p_0 and its density ρ_0, and to move without rotation through a nozzle having the shape defined in figure 1 and a size defined by its throat width $2T$; the position of any point in the field of flow was

---

† The reason for the difference in shape is explained in § 11.

defined by 'non-dimensional' (i.e. purely numerical) co-ordinates x and y, these standing for its true (Cartesian) co-ordinates expressed as multiples of T; and the total mass flow through the nozzle was denoted by $2M$. By conformal transformation of the field of flow into a rectangle (§§ 6–8*)[†] a new system of numerical co-ordinates α, β was devised such that $\beta = 0$ on the centre-line, $\beta = \text{const.} = \beta_1$ (say) at the nozzle wall on the side of y positive (figures 2)[‡].

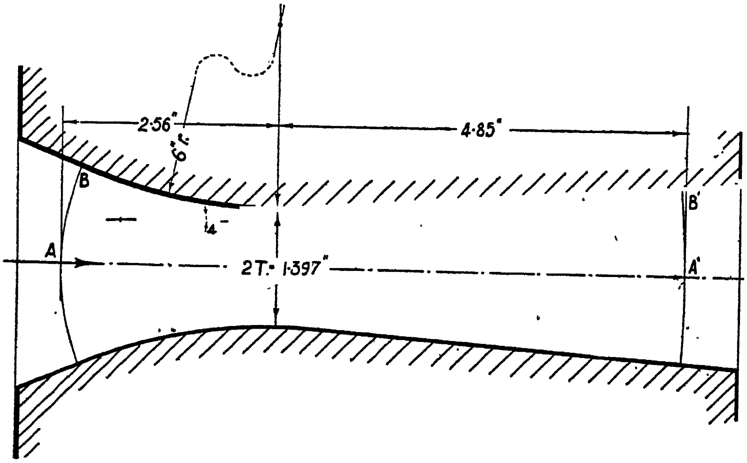


FIGURE 1

On this understanding, $M\psi$ standing for the 'mass-flow function' (so that ψ is purely numerical), ψ has to satisfy the boundary conditions

$$\left.\begin{aligned}\psi &= 0 \quad \text{when} \quad \beta = 0,\\ &= 1 \quad \text{when} \quad \beta = \beta_1,\end{aligned}\right\} \quad (20)^* \qquad (1)$$

and its governing equation was shown to be

$$\nabla^2(\chi.\psi) - \psi.\nabla^2\chi = 0, \quad (19)\text{A}^* \qquad (2)$$

where

$$\chi^2 = F\left[\mu^2 h^2\left\{\left(\frac{\partial\psi}{\partial\alpha}\right)^2 + \left(\frac{\partial\psi}{\partial\beta}\right)^2\right\}\right] \quad (18)\text{A}^* \qquad (3)$$

and the form of the function F is known. In (2)

$$\nabla^2 \equiv \frac{\partial^2}{\partial\alpha^2} + \frac{\partial^2}{\partial\beta^2} \quad (21)^* \qquad (4)$$

and $\chi^{-2} = \rho$ denotes the local density expressed as a multiple of ρ_0; (13)* (5)

and in (3) $$h^2 = \left|\frac{d(\alpha+i\beta)}{d(x+iy)}\right|^2 = \left(\frac{\partial\alpha}{\partial x}\right)^2 + \left(\frac{\partial\alpha}{\partial y}\right)^2 = \left(\frac{\partial\beta}{\partial x}\right)^2 + \left(\frac{\partial\beta}{\partial y}\right)^2, \quad (16)^* \qquad (6)$$

[†] Asterisks indicate section and equation numbers of the earlier paper.

[‡] According to (i) of § 12, $\partial\alpha/\partial x = -\partial\beta/\partial y$. In figures 7*, since it is convenient to have $\partial\alpha/\partial x$ and $\partial\beta/\partial y$ both positive, purely arbitrary values *increasing from left to right* were attached to α.

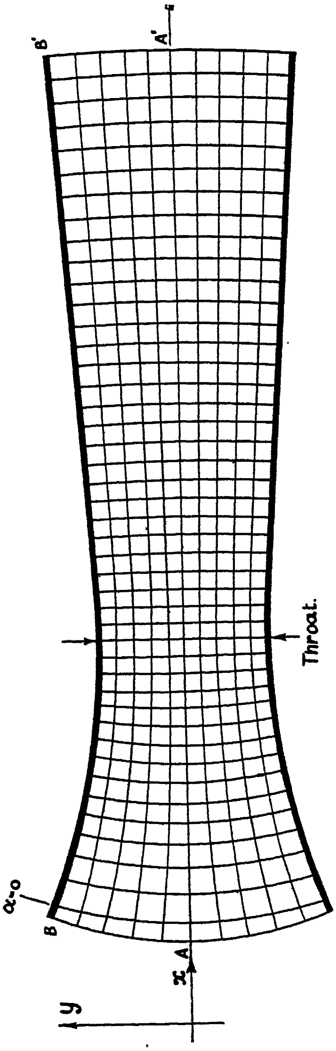


FIGURE 2*a*

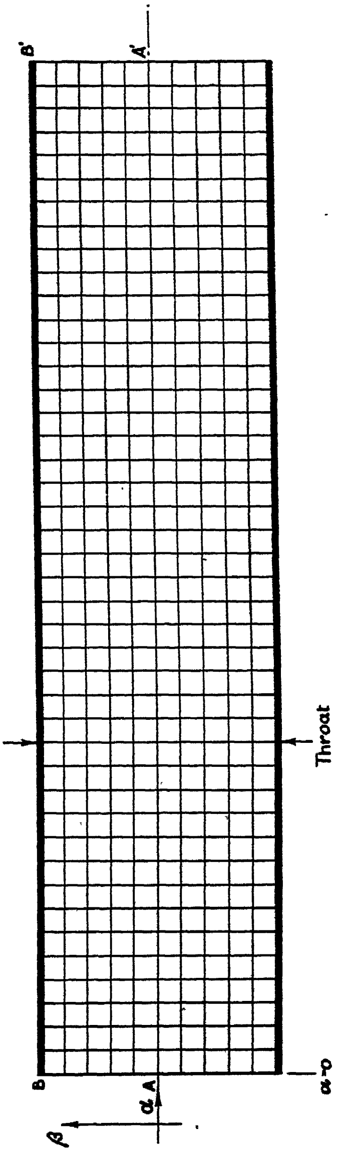


FIGURE 2*b*

i.e. h is the modulus of the conformal transformation. The quantity

$$\mu = M/\rho_0 a_0 T, \quad (14)^*$$

where a_0 stands for the velocity of sound in the gas at starting, so that

$$a_0^2 = \gamma p_0/\rho_0. \tag{7}$$

With this notation (differing from that of the earlier paper in that the dashes there used to distinguish 'non-dimensional' quantities are here suppressed) the component velocities at any point in the fluid are u and v where

$$\frac{u}{a_0} = \mu\chi^2\frac{\partial\psi}{\partial y}, \quad -\frac{v}{a_0} = \mu\chi^2\frac{\partial\psi}{\partial x}, \quad (2)^*, (13)^* \tag{8}$$

therefore $$q^2 = \mu^2\chi^4\left\{\left(\frac{\partial\psi}{\partial x}\right)^2 + \left(\frac{\partial\psi}{\partial y}\right)^2\right\},$$

i.e., by (5) and (6), $$\rho^2q^2 = \mu^2h^2\left\{\left(\frac{\partial\psi}{\partial\alpha}\right)^2 + \left(\frac{\partial\psi}{\partial\beta}\right)^2\right\} \quad (22)^* \tag{9}$$

when q denotes the local velocity of the gas *expressed as a multiple of* a_0. Equation (3) can now be written in the form

$$\frac{1}{\rho} = \chi^2 = F(\rho^2q^2),$$

the form of F being calculable from the properties of the gas. Figure 3 shows ρq and q $(= \chi^2\rho q)$ plotted against χ, for air assumed (§ 10*) to start from rest at temperature 15° C and pressure 100 lb./sq. in. (absolute). It was plotted from the relation

$$\rho^2q^2 = 5\chi^{-4}(1-\chi^{-0\cdot8}), \quad (23)^* \tag{10}$$

which in § 10* was derived, for these starting conditions, on the assumption that air can be treated as a perfect gas $(\gamma = 1\cdot4)$, and which on that understanding may be used in place of (3).

5. The reason for the introduction of conformal transformation was that by using α and β as independent variables, in place of x and y, the boundary is transformed into a rectangle, consequently relaxation can be effected on a square net (figure 2*b*) having all its meshes complete, i.e. with no 'irregular stars'. The cost of this simplification is the labour entailed in the determination of h, of which the values at nodal points must be known with accuracy in order that χ-values may be accurately determinable from (9) and (10). Three-figure accuracy was claimed (§ 15*) for Green's computed values, and in his figure 7*c** h was recorded to five figures, which it was claimed (cf. footnote to § 15*) satisfy to this accuracy the requirement that $\log h$ must be plane-harmonic. We shall show (§ 11) that his claims are not entirely justified; but it should be emphasized that the fact had no importance in his subsonic investigations, for which three-figure accuracy in h was sufficient.

6. This paper does *not* make use of relaxation methods, but of the iterative ('strip integration') method which was devised (cf. § 19*) to meet the failure of those methods as applied to the *supersonic* regime. We now summarize §§ 11–12* and 22–24*.

Multiplied through by χ, equation (2) can be put into the form

$$\frac{\partial}{\partial\beta}\left(\chi^2\frac{\partial\psi}{\partial\beta}\right) = -\frac{\partial}{\partial\alpha}\left(\chi^2\frac{\partial\psi}{\partial\alpha}\right). \quad (37)^* \qquad (11)$$

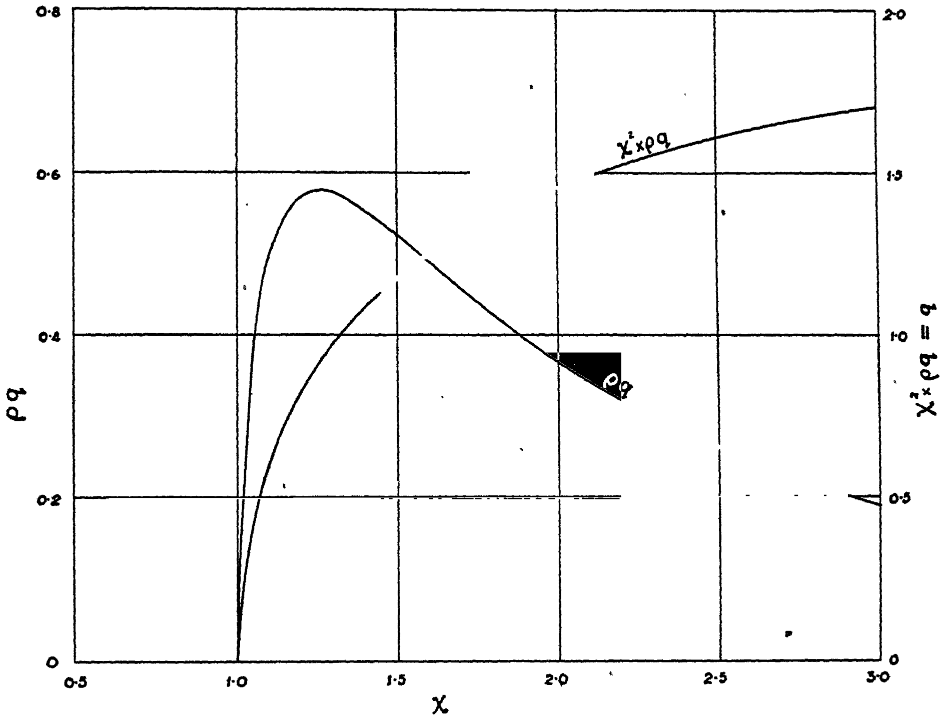


FIGURE 3

Now we may expect that contours of ψ, in the x-y plane, will approximate closely to contours of β; and this means that the β-differentials of ψ will predominate in (11), so that without serious error we may suppress the α-differentials to obtain

$$\left.\begin{aligned} &\frac{\partial}{\partial\beta}\left(\chi^2\frac{\partial\psi}{\partial\beta}\right) = 0, \\ \text{i.e.}\qquad &\chi^2\frac{\partial\psi}{\partial\beta} = \text{const.} = A \text{ (say).}\end{aligned}\right\} \quad (24)^* \qquad (12)$$

Equally $\partial\psi/\partial\beta$ will predominate over $\partial\psi/\partial\alpha$ in the expression (9), so that without serious error we may suppress the α-differential to obtain

$$\mu h\frac{\partial\psi}{\partial\beta} = \rho q. \quad (25)^* \qquad (13)$$

(Here, in fact, the error is of second order when $(\partial\psi/\partial\alpha)/(\partial\psi/\partial\beta)$ is small.) Now eliminating $\partial\psi/\partial\beta$ between (12) and (13), we have *approximately*

$$q = \rho q \chi^2 = A\mu h, \quad (26)^* \tag{14}$$

where, by the second of (12)

$$\psi = \int^{\beta} \frac{\partial\psi}{\partial\beta'} d\beta' = A \int^{\beta} \chi^{-2} d\beta'. \quad (27)^* \tag{15}$$

Applied to (15), the boundary condition (1) gives

$$\left.\begin{aligned} 1 &= A \int_0^{\beta_1} \chi^{-2} d\beta, \\ \text{i.e.} \qquad \mu &= \mu A \int_0^{\beta_1} \chi^{-2} d\beta. \end{aligned}\right\} \quad (28)^* \tag{16}$$

For our assumed starting conditions, in virtue of (10), (14) may be replaced by

$$\left.\begin{aligned} A\mu h = q = \rho q \chi^2 &= \sqrt{\{5(1-\chi^{-0\cdot 8})\}}, \\ \text{giving} \qquad \chi^{-2} &= \left(1 - \frac{A^2\mu^2h^2}{5}\right)^{2\cdot 5}, \\ \text{and hence, according to (16),} \\ 1 &= A \int_0^{\beta_1} \left(1 - \frac{A^2\mu^2h^2}{5}\right)^{2\cdot 5} d\beta. \quad (36)^* \end{aligned}\right\} \tag{17}$$

It should be remarked that the range of q is always limited. Thus $0 < q < \sqrt{5}$ according to (17).

7. For every α-line of the rectangular field (§ 5) we have (when the conformal transformation has been effected) a known distribution of h. Therefore, for any assumed value $A\mu$ such that $A\mu h$ does not fall outside the range of q, we can deduce the distribution of q and thence, using (17), of χ. Then, evaluating the definite integral in the second of (16) and inserting the assumed value of $A\mu$, we have a corresponding value for μ, so can construct a curve of the type of figure 4. $\mu = 0$ at both ends of the range of $A\mu$, and it has a maximum value ($\mu_{\text{max.}}$ say) at some intermediate point.

8. All of this has been based on the assumption (§ 6) that the exact equation (11) may be replaced by the first of (12). For closer approximation, let the solution found as above (ψ_1, χ_1, say) be substituted on the right of (11), and let a second approximation (say, ψ_2, χ_2) be found from the resulting equation

$$\frac{\partial}{\partial\beta}\left(\chi^2 \frac{\partial\psi}{\partial\beta}\right) = -\frac{\partial}{\partial\alpha}\left(\chi_1^2 \frac{\partial\psi_1}{\partial\alpha}\right). \quad (37)^* \tag{18}$$

Thereby we obtain, in place of the second of (12),

$$\chi^2 \frac{\partial \psi}{\partial \beta} = A - F_1(\beta), \quad (38)^* \tag{19}$$

A as before denoting a constant of integration, and where

$$F_1(\beta) = \int_0^\beta \frac{\partial}{\partial \alpha}\left(\chi_1^2 \frac{\partial \psi_1}{\partial \alpha}\right) d\beta' \quad (41)^* \tag{20}$$

is calculable. *The expression* (13) *holds as before*, and on eliminating ψ between it and (19) we have, in place of (14),

$$q = \rho q \chi^2 = A\mu h - \mu h F_1(\beta). \quad (39)^* \tag{21}$$

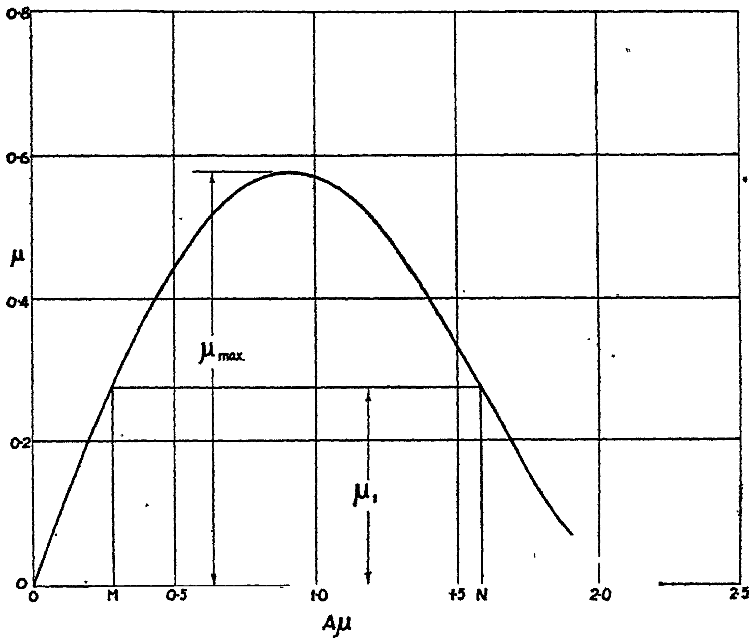


FIGURE 4

$$\left.\begin{aligned}
&\text{Also from (19) we have, in place of (15),}\\
&\qquad \psi = A\int_0^\beta \chi^{-2}\, d\beta' - \int_0^\beta \chi^{-2} F_1(\beta')\, d\beta',\\
&\text{and hence, in place of the second of (16),}\\
&\qquad \mu = \mu A \int_0^{\beta_1} \chi^{-2}\, d\beta - \mu \int_0^{\beta_1} \chi^{-2} F_1(\beta)\, d\beta. \quad (40)^*
\end{aligned}\right\} \quad (22)$$

The procedure of § 7 is now no longer applicable, because q is not determinable from (21) unless μh as well as $A\mu h$ is known. But our first approximation was derived on an assumption which makes $F(\beta)$ small, and on that account the new value of μ which corresponds with any assumed value of A will differ but little from μ_1, its value according to the first approximation: hence we may with sufficient accuracy replace (21) and the second of (22) by

$$\left.\begin{aligned} & q = \rho q\chi^2 = A\mu h - \mu_1 h F_1(\beta) \\ \text{and by} \qquad & \mu = A\mu \int_0^{\beta_1} \chi^{-2}\, d\beta - \mu_1 \int_0^{\beta_1} \chi^{-2} F_1(\beta)\, d\beta, \end{aligned}\right\} \quad (42)^* \qquad (23)$$

and then (μ_1 and $F_1(\beta)$ being known) the construction of figure 4 can proceed as before. For our assumed starting conditions, i.e. when (10) is satisfied, the first of (22) becomes

$$\left.\begin{aligned} & \psi = \int_0^{\beta} \left(1 - \frac{q^2}{5}\right)^{2\cdot 5} \{A - F_1(\beta')\}\, d\beta', \\ & \text{and the second of (22) becomes} \\ & 1 = \int_0^{\beta_1} \left(1 - \frac{q^2}{5}\right)^{2\cdot 5} \{A - F_1(\beta)\}\, d\beta, \\ & q \text{ being given by the first of (23).} \end{aligned}\right\} \qquad (24)$$

Application of the iterative method to determine the supersonic regime

9. When the *supersonic* regime is in question, μ must be given its limiting value $\mu_{\text{crit.}}$, for which Osborne Reynolds's treatment yields the approximation

$$(5/6)^3 = 0\cdot 578{,}704 \quad (\S\,10^*).$$

For reasons stated in § 24* it is of great importance that this value (which is *not* known in advance) be estimated closely, and clearly, for that purpose, similar accuracy is needed in the evaluation of h. Whereas three-figure accuracy sufficed (§ 5) for Green's investigation of speeds below the speed of sound, here we shall need the five-figure accuracy which was claimed for the values recorded in his figure 7c*.

The method used to obtain those values was indicated in a footnote to § 15*. Having effected the conformal transformation (§ 5) and deduced three-figure values of h by means of (6), Green accepted these values as having five-figure accuracy *on the boundary* ($\beta = \beta_1$), and deduced for every internal point a corresponding value of $\log h$, which is plane-harmonic. The underlying idea was that these values, although they might not correspond (to five-figure accuracy) with the form of nozzle shown in figure 1, *would nevertheless imply a form so near as to be identical within the limits of accuracy that are attainable in manufacture*; and on this account no attempt was made to determine the exact form that they implied.

This was sound procedure as concerned the computation of subsonic flow patterns, but something more is needed when (as in this paper) we wish to compare with Reynolds's estimate the value given by our iterative method for the critical mass flow.

If in fact our analysis implies a throat width not of $2T$ but of some other value $2\mathbf{T}$, then the estimate of $\mu_{\text{crit.}}$ according to Reynolds's theory is not $(5/6)^3$, as stated above, but (cf. § 10*)

$$\mu_{\text{crit.}} = M_{\text{crit.}}/(\rho_0 a_0 T) = (5/6)^3\,\mathbf{T}/T. \tag{25}$$

So, for an accurate comparison, $\mathbf{T}/T$ must be accurately computed.

10. This is a simpler matter than a derivation of β-values everywhere (§ 12). For according to (6), if $\partial\beta/\partial x \doteqdot 0$ as is known (from Green's figure 7b*) to be true near the throat, then to a close approximation

$$h = \frac{\partial\beta}{\partial y},$$

and so

$$\mathbf{T} = T\int_0^{\beta_1}\frac{\partial y}{\partial\beta}\,d\beta = T\int_0^{\beta_1} h^{-1}\,d\beta. \tag{26}$$

The distribution of h

11. Here again our accuracy will depend on the accuracy of the accepted h-values. Therefore we are obliged to notice one or two slight errors in Green's figure 7c*, which would give trouble in a computation of the supersonic regime, although for his purpose they are unimportant.

The boundary values which Green assumed for h (§ 9) do not (to five-figure accuracy) vary smoothly in every part of the nozzle wall: by what must have been a slip in computation, a value 100,800 was there given to h at a point where 100,000 is indicated by a plotting of h against α, and thus a local 'disturbance' was introduced. This, of course, entailed consequential errors in neighbouring h-values as determined by the requirement that $\log h$ must be plane-harmonic.

Having reason (cf. § 2) to believe that the supersonic regime is in fact unstable, we sought to avoid all features in the assumed shape of nozzle which might militate against the finding of this regime by computation, and accordingly we decided not to retain Green's shape exactly, but to modify his assumed h-values by a preliminary smoothing process *before* proceeding to deduce internal values from the consideration that $\log h$ is plane-harmonic. Figures 5 record the resulting h-values, which compare with those of his figure 7c*. To the extent that they are different, they obviously relate to a different form of nozzle; and the problem is thus presented, of determining this new form.

The nozzle shape

12. The conformal transformation determines $\alpha + i\beta$ as a function of $x + iy$, and it therefore follows (in the theory of functions of a complex variable) that the differential coefficient $d(\alpha+i\beta)/d(x+iy)$ is a function of either. Moreover,

$$\frac{d(\alpha+i\beta)}{d(x+iy)} = \frac{\partial\alpha}{\partial x} + i\frac{\partial\beta}{\partial x} = -\frac{\partial\beta}{\partial y} + i\frac{\partial\beta}{\partial x} \tag{i}$$

$$= -h(\cos\theta + i\sin\theta) = -he^{i\theta}, \tag{ii}$$

where h is the quantity defined in (6) of § 4, and where θ, *measured counter-clockwise* from the direction of x, is the angle made with that direction by the contour $\beta = \text{const.}$, so that

$$0 = \cos\theta\frac{\partial\beta}{\partial x} + \sin\theta\frac{\partial\beta}{\partial y}, \quad h = \frac{\partial\beta}{\partial s} = \cos\theta\frac{\partial\beta}{\partial y} - \sin\theta\frac{\partial\beta}{\partial x}. \tag{iii}$$

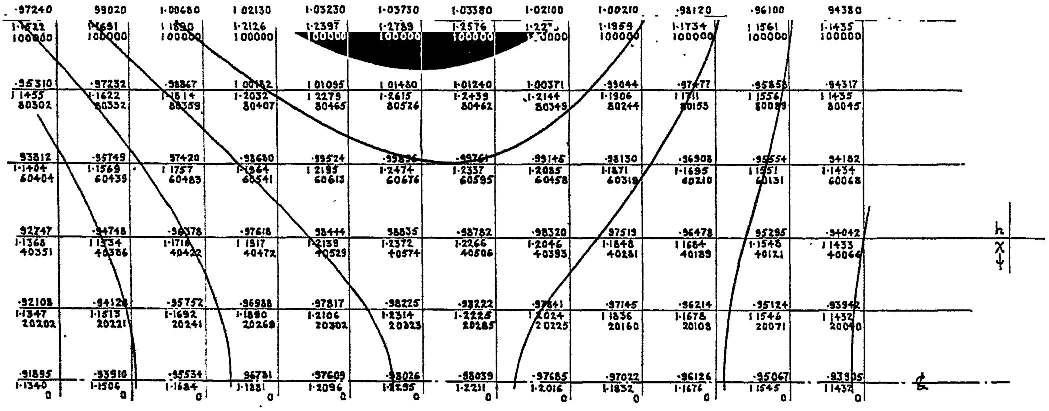


FIGURE 5*a*. h, χ and ψ for limiting subsonic regime, with contours of constant h.

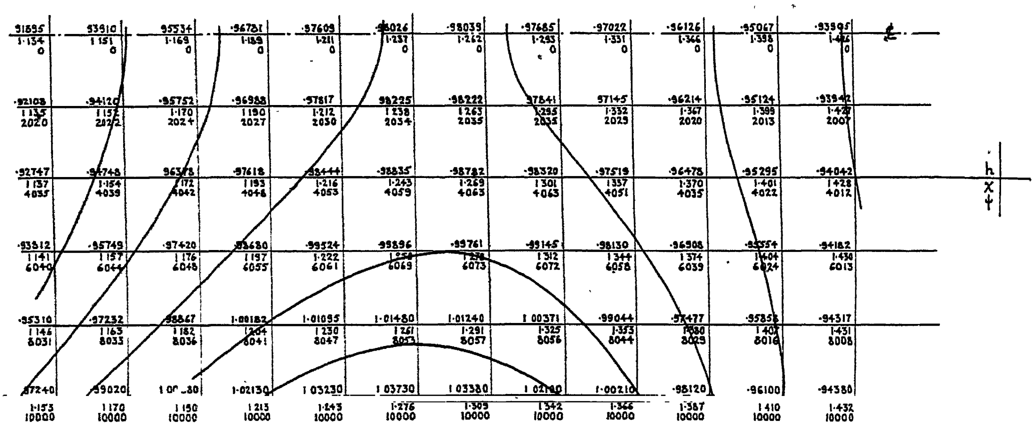


FIGURE 5*b*. h, χ and ψ for supersonic regime, with contours of constant h.

Accordingly $he^{i\theta}$, and therefore $(\log h + i\theta)$, are functions of $x+iy$ and of $\alpha+i\beta$. Consequently

$$\frac{\partial\theta}{\partial\alpha} = \frac{\partial}{\partial\beta}(\log h), \quad -\frac{\partial\theta}{\partial\beta} = \frac{\partial}{\partial\alpha}(\log h), \tag{27}$$

and from these relations, given $\log h$, we can determine θ (its conjugate) by the methods of Part V (Gandy & Southwell 1940). Then, finally, x and y may be deduced by integration from

$$-\frac{\partial x}{\partial\alpha} = \frac{\partial y}{\partial\beta} = \frac{\cos\theta}{h}, \quad \frac{\partial x}{\partial\beta} = \frac{\partial y}{\partial\alpha} = -\frac{\sin\theta}{h}, \tag{28}$$

which are consequences of (i) and (ii). (Where θ is small, so that $\cos\theta \doteqdot 1$, the first of (28) implies that

$$h = \frac{\partial\beta}{\partial y}$$

as assumed in § 10.)

The form of nozzle thus derived from the h-values recorded in figures 5 is practically indistinguishable from Green's form (figure 1). To fix the value of T/T, L.F. computed the integral $\int_0^{\beta_1} h^{-1}\,d\beta$ (§ 10) for three sections close to the section marked 'throat' in figures 2. Table 1 records his results, from which, by normal methods of interpolation, it was deduced (i) that $\alpha = 28{\cdot}56 \times 10^4$ for the true throat section, where T/T as given by (26) has its minimum value; (ii) that this minimum value is 1·00189.

TABLE 1

$\alpha \times 10^{-4} =$	26	28 ('throat')	30
$\mathrm{T}/T =$	1·00612	1·00209	1·00324

Substituting this value for T/T in (25), we have according to Osborne Reynolds's approximate treatment

$$\mu_{\text{crit.}} = (5/6)^3 \times 1{\cdot}00189 = 0{\cdot}57980. \tag{29}$$

II. RESULTS OF COMPUTATION BY THE ITERATIVE METHOD

13. With h-values as recorded in figures 5, computations on the lines of §§ 6–8 were made (again, by L.F. with assistance from Miss Gillian Vaisey) to determine the value of $\mu_{\text{crit.}}$. The means to this determination were outlined in § 24* of the earlier paper, which we now summarize.

For every 'section' (i.e. contour of α in figure 5*b*), the iterative method can be used to construct a curve of the type of figure 4, in which μ ranges from 0 to some maximum value $\mu_{\text{max.}}$. Only values less than $\mu_{\text{max.}}$ are admissible, and so (since in any possible regime μ will have the same value for every 'section') the wanted (limiting) value of μ is the least $\mu_{\text{max.}}$ which is discoverable. Clearly it will be the $\mu_{\text{max.}}$ for some section near the throat, but it is not certain that this section will be one of those that we investigate. However, having values of $\mu_{\text{max.}}$ for three or four values of α near the throat, it is an easy matter to deduce (by customary interpolation methods) the α-value for which $\mu_{\text{max.}}$ has its minimum value, and thence to compute this minimum. Then, accepting the computed value and reverting to the curves (of type figure 4) which have been constructed for other sections, it is again an easy matter to deduce for each of them a value of A (§ 6), and with this to compute ψ for the relevant section from (15).

Two such values of A will be admissible, according as we are concerned with a regime of the type of (*a*) or (*b*), § 1, i.e. with a subsonic or supersonic regime. Thus in figure 4 (on its scale of $A\mu$)

$$OM = \mu\,A_{\text{sub.}}, \quad ON = \mu\,A_{\text{sup.}},$$

OM entailing subsonic and ON supersonic velocities at the 'section' considered. Upstream of the throat section we must postulate subsonic conditions: downstream we may give to A (at every section) its appropriate value either of $A_{\text{sub.}}$ or $A_{\text{sup.}}$.

Estimation of $\mu_{\text{crit.}}$ *for subsonic regimes*

14. Proceeding on these lines, L.F. first calculated to a first approximation (§ 7) values of $\mu_{\text{max.}}$ for the section marked as 'throat' in figures 2, and for the adjoining section on either side. His results are recorded in table 2: from them, by customary methods, he deduced

$$\mu_{\text{crit.}} = 0{\cdot}579{,}60 \tag{i}$$

as a first approximation to the critical mass flow, and established that the true throat section lies slightly downstream of the central section in table 2 ($\alpha = 28{\cdot}5_5 \times 10^4$).

TABLE 2

$\alpha \times 10^{-4} =$	26	28 ('throat')	30
$\mu_{\text{max.}}$ (1st approx.)	0·582,11	0·579,71	0·580,41
$\mu_{\text{max.}}$ (2nd approx.)	0·582,07	0·579,43	0·580,30

He then proceeded to a second and third approximation in the manner of § 8. Giving to μ his first approximation (i) to $\mu_{\text{crit.}}$, he deduced second approximations to ψ and to χ for the five central sections, and thence $F(\beta)$ for the three central sections; then (giving to μ_1, in (24), the same value 0·579,60) he could obtain new estimates of $\mu_{\text{max.}}$ for the three central sections, as recorded in the last line of table 2. From these a second approximation to the critical mass flow could be deduced as before, viz.

$$\mu_{\text{crit.}} = 0{\cdot}579{,}32. \tag{30}$$

The whole process was then repeated with this new value 0·579,32, instead of 0·579,60, given to μ and to μ_1. No change resulted in the estimate of $\mu_{\text{crit.}}$, and accordingly (30) was accepted.

In § 10* of the earlier paper it was shown that Osborne Reynolds's estimate (29) must be regarded as an upper limit to the correct value, i.e. (since it can hardly be exact) as an *overestimate*. This expectation is realized, since our result (30) is 0·083 % less.

15. Next, giving to μ its critical value as recorded in (30), L.F. proceeded to determine (in accordance with § 13) a value of $A_{\text{sub.}}$ for every section in the range of α which is covered by figures 5. Outside that range new computations were not needed, for the reason that *h is sensibly independent of β.*

(On a section where h is constant we have, from (14),

$$q = A\mu h = \text{constant},$$

therefore $$\chi = \text{const., according to (10).}$$

Consequently (15) and the first of (16) simplify to

$$\psi = A\chi^{-2}\beta, \quad 1 = A\chi^{-2}\beta_1,$$

whence

$$\psi = \beta/\beta_1, \tag{i}$$

as we should expect. Now replacing A by χ^2/β_1, we have from (14)

$$\beta_1 q = \mu h \chi^2, \tag{ii}$$

in which μ, h, β_1 are known. Consequently q and χ can be determined separately from a diagram of the type of figure 4 or (in our particular problem) from the equation

$$5(1-\chi^{-0\cdot 8}) = q^2 = (\mu h \chi^2/\beta_1)^2, \tag{31}$$

which comes from (10) and (ii).

All of this is based upon the simple theory of §§ 6–7. But no correction comes from the amended theory of § 8, because when ψ_1 is given by (i) then $\partial\psi_1/\partial\alpha = 0$ and so $F_1(\beta) = 0$.)

Determination of the critical subsonic regime

16. The resulting 'subsonic' regime is exhibited in figure 6, which compares with figures 8–11 of the earlier paper. Results were found to converge rapidly, no more than three cycles of the iterative process being required at any section.

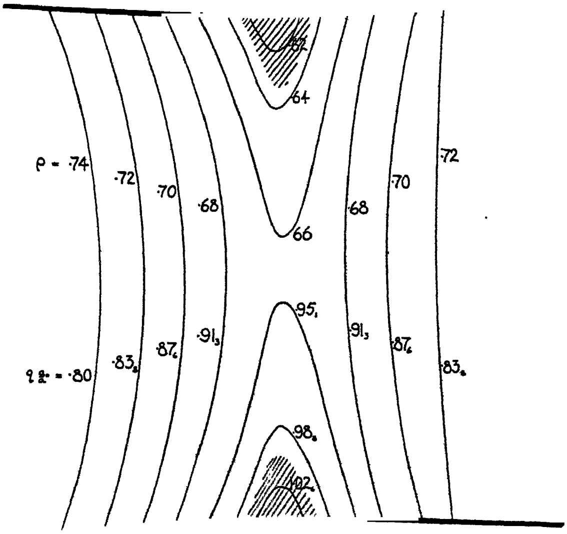


Figure 6. Limiting subsonic regime in the region of the 'throat'.

A feature of the solution is the attainment of velocities exceeding the speed of sound in a localized region (shaded in figure 6) near the walls of the throat section. This has importance in view of a remark by von Kármán (1941): 'It appears that in the case of flows that are partially subsonic, partially supersonic, the existence of continuous velocity fields is one of the questions of basic importance.'

Estimation of $\mu_{\text{crit.}}$ for the supersonic regime

17. We turn to the more difficult problem of the supersonic regime, which on our assumptions will be (as in Osborne Reynolds's simple treatment) *unique*. (In practice, of course, we must contemplate that our nozzle discharges into a reservoir maintained at *any* constant pressure, and a question thus arises as to what happens if this pressure is different from what our unique solution requires. That question, however, arises also in relation to Reynolds's treatment,† and it introduces physical considerations with which a computational paper has no concern.)

The first point to be noticed is that $\mu_{\text{crit.}}$, determined in the manner of § 14, has a different value according as the velocity is subsonic or supersonic downstream of the throat. The simple treatment of § 7, of course, makes the subsonic and supersonic values identical, and a like result (with a slight difference in the numerical value) follows from Reynolds's treatment; but the quantity $F_1(\beta)$, which enters into the more exact theory of § 8, is different in the two regimes, and this difference is reflected in the value of $\mu_{\text{crit.}}$. Proceeding in the manner of § 14, L.F. obtained supersonic values as recorded in table 3, which compares with table 2 for the subsonic regime. The throat (defined as that 'section' for which $\mu_{\text{max.}}$ has its least value) was found to be given by

$$\alpha = 28{\cdot}73 \times 10^4, \tag{32}$$

i.e. to be slightly further downstream than before; and the accepted estimate of the critical mass flow was that given by

$$\mu_{\text{crit.}} = 0{\cdot}579{,}58, \tag{33}$$

which is higher by 0·045 % than the subsonic value (30).

TABLE 3

$\alpha \times 10^{-4}$	26	28	30
$\mu_{\text{max.}}$ (1st approx.)	0·582,11	0·579,71	0·580,41
$\mu_{\text{max.}}$ (2nd approx.)	0·582,03	0·579,68	0·580,37

Determination of the critical supersonic regime

18. Accepting this value, L.F. proceeded to compute for all 'sections' of figure 5*b* the supersonic distributions of ψ and χ, with results which are shown in figure 7. Difficulties were confronted which had not been met in our subsonic computations, in that the iterative process was found to *diverge*. There appeared to be a non-zero lower limit to the algebraic sum of the nodal values of $F_1(\beta)$, which limit was attained, in general, after two cycles of the process: thereafter the sum increased, much in the way that 'residual forces' increase when normal relaxation methods are employed on *eigenwerte* problems with an incorrect value for the characteristic number. In our accepted results, nodal values of $F_1(\beta)$ were 'stable' (i.e. did not alter in successive

---

† Cf. von Kármán 1941, p. 340.

stages) to one part in a thousand of the values of ψ and χ, and these had like accuracy in the sense that a change of that order made matters worse. Higher accuracy seemed unattainable, and the impression was gained that although the boundary h's had been 'smoothed' as far as was practicable, discontinuities remaining in the implied nozzle boundary might be the cause of disturbances large enough to prevent the finding of an unstable regime.

The subsonic regime appeared to be, in comparison, highly stable. Quite large $F_1(\beta)$'s were manageable, and values of opposite sign on adjacent strips tended to ultimate 'stability', whereas the opposite was found in the supersonic solution. Our subsonic solution (figure 6) is believed to be accurate within one part in ten thousand.

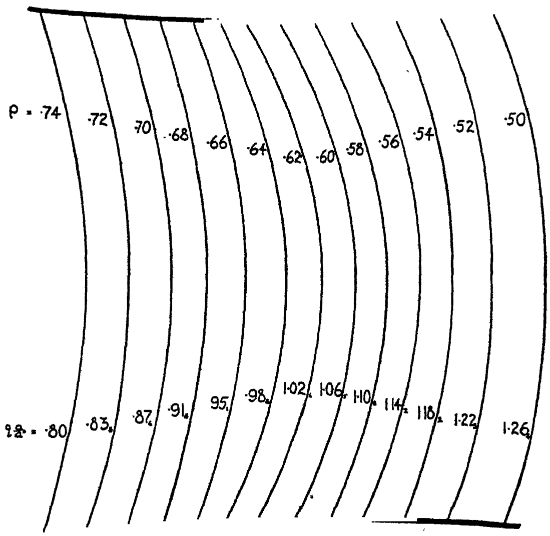


Figure 7. Supersonic regime in the region of the 'throat'.

Conclusion

19. Our conclusions are as stated in § 3, with the addition that irrotational flow (so far as can be judged from the convergence of our iterative process) is stable from low velocities up to those entailed in our 'limiting *subsonic* regime', but is unstable (though possible, according to the governing equations) in the unique *supersonic* regime.

It may now be said with confidence that *any* shape of nozzle can be treated in two dimensions, and with slight modification our method could be applied to nozzles having axial symmetry. The labour entailed would of course be somewhat greater.

Grateful acknowledgment is made of grants from the Ministry of Aircraft Production and from the Department of Scientific and Industrial Research which permitted, respectively, L. F. to make and Miss Vaisey to assist in the computations here described.

REFERENCES

Gandy, R. W. G. & Southwell, R. V. 1940 ('Part V') *Phil. Trans.* A, **238**, 453.
Green, J. R. & Southwell, R. V. 1944 ('Part IX') *Phil. Trans.* A, **239**, 367.
von Kármán 1941 *J. Aero. Sci.* **8**, 337.
Reynolds, O. 1886 *Phil. Mag.* **21**, 185.
Taylor, G. I. & Sharman, C. F. 1928 *Proc. Roy. Soc.* A, **121**, 194.

The production of cosmic-ray bursts by mesotrons

BY C. B. O. MOHR, PH.D., *Lecturer in Physics*, AND G. H. STAFFORD, M.SC.,
University of Cape Town

(*Communicated by H. S. W. Massey, F.R.S.—Received* 26 *August* 1943)

The frequency of occurrence of cosmic-ray bursts under large thicknesses of iron and lead has been investigated as a function of the thickness. A definite maximum is found to occur in the curves for the smaller bursts, its prominence diminishing with bursts of increasing size.

The dependence of the burst frequency on the size of the ionization chamber and the proximity of the shielding material to the chamber has also been investigated. The results suggest the production of showers of mesotrons which spread out, and ultimately give rise to electron showers originating in widely separated areas in the material above, and round the sides of, the ionization chamber.

It is now well known that the cosmic radiation reaching the earth's surface comprises two main ionizing components, a soft component consisting of fast electrons and a hard component consisting of mesotrons.

The fast electron component is absorbed in a few centimetres of heavy material as the result of a cascade process in which each fast electron gives rise to a shower of soft electrons. The frequency of showers of a given size rises to a maximum and then falls again as the thickness of the material is increased, giving the familiar 'Rossi curve'. The processes involved have been fairly extensively investigated with various arrangements of coincidence counters, and with ionization chambers which record bursts of ionization produced by showers, and the main features of the observations have been satisfactorily accounted for by the quantum theory (Froman & Stearns 1938).

The Rossi curve, after an initial fall beyond the maximum, flattens out and then falls off much more slowly, indicating the presence of a harder component, the mesotron component, capable of penetrating several metres of heavy material. The occurrence of large cosmic-ray bursts under great thicknesses of matter has been ascribed to the following process (Christy & Kusaka 1941). The mesotrons produce fast electrons, either (*a*) in head-on collisions with atomic electrons, or (*b*) through the intermediary of hard quanta produced by retardation of the mesotrons in nuclear collisions ('Bremsstrahlung'), these quanta subsequently giving rise to fast electron pairs. The fast electrons then give rise to showers of slower electrons by cascade.

It was thought for some time that for these larger thicknesses, corresponding to the flatter part of the Rossi curve, equilibrium between mesotrons and electrons was reached, and that the showers of electrons formed did not extend over large areas. Certain observers have, however, obtained indications of the occurrence of more complicated processes than those mentioned above. Some, using coincidence counters, have found traces of a second maximum in the Rossi curve at about 30 cm. of iron and about 17 cm. of lead, though others have been unable to confirm this result (for a list of references see Nereson 1942). Swann & Ramsay (1940*a*), using large numbers of self-recording counters arranged in layers, one under the other, have determined the rate of occurrence of two-particle showers under different thicknesses of tin, and found a second maximum in their curve. Broussard & Graves (1941), using an expansion chamber, have investigated the frequency of occurrence of two-particle showers in lead, and obtained a maximum frequency under 17 cm.

The suggestion (Bhabha 1938; Heitler 1938) that the second maximum in the Rossi curve is due to the production of mesotrons in the material receives support from more recent calculations (Heitler 1941; Wilson 1941; Heitler & Peng 1942), which have shown that there is an appreciable probability of the production of mesotron showers by sufficiently energetic primary mesotrons. That such multiple processes involving mesotron creation do occur is now confirmed by a number of experiments (see Discussion).

We have observed the frequency of occurrence of bursts under thicknesses of iron ranging from 14 to 90 cm., and of lead ranging from 10 to 40 cm., and found a pronounced maximum which, it would seem, arises from the production of mesotron showers. We have also observed the burst frequency using two ionization chambers of different sizes as well as different arrangements of the larger chamber and shielding material, and obtained somewhat unexpected results. A mechanism is proposed for the explanation of these results as well, in terms of the production of mesotron showers in the material.

APPARATUS

Most of the measurements were carried out with a large cylindrical steel ionization chamber of wall thickness 0·3 cm. containing air at a total pressure of 7·2 atm. The electrodes were in the form of concentric cylinders of length 57·5 cm., and diameter 22·3 and 5·0 cm. respectively, and consisted of two iron wire grids soldered to iron-rod frameworks with zinc to minimize the 'natural effect'. The chamber was assembled and used in a laboratory which has always been free of radioactive contamination. Some measurements were also taken with a small duralumin chamber of wall thickness 1 cm. containing air at a total pressure of 90 atm. The electrodes, of spun aluminium, were approximately in the form of concentric cylinders of mean diameter 10·0 and 4·0 cm. respectively with hemispherical ends.

For the first runs the high potential for the electrodes was obtained from 704 small dry cells, arranged in eleven boxes placed one above the other and carefully insulated from each other. After some weeks of hot weather these batteries began to dry out, and gave fluctuations of voltage which, though relatively small, were for our purpose hopelessly big. It has previously been pointed out (Carmichael 1936) that in this work sudden fluctuations in the potential of more than 1 part in 10^6 must be avoided. The dry batteries were replaced by 30 new Exide 10 V high-tension units, and these have continued to give the necessary steadiness. While the smaller potential resulted in a longer time of collection of the ions, this should not affect the variation of the burst frequency with thickness of material above the chamber.

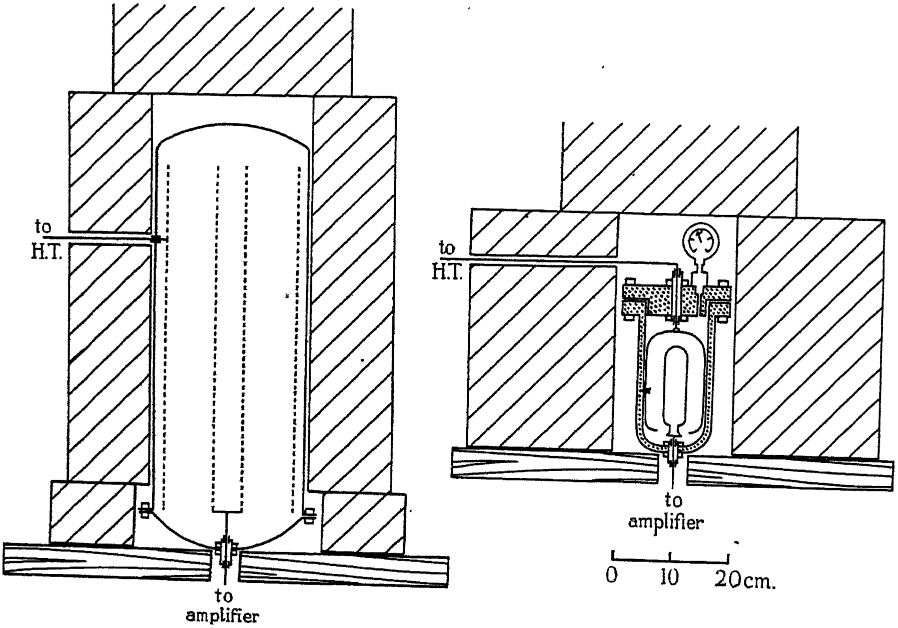


FIGURE 1. Geometrical arrangement of the chambers with iron shielding.

The inner cylindrical electrode was connected directly to the control grid of an electrometer valve set up in a balanced circuit (du Bridge 1933), and changes in the steady plate current were detected with a Moll galvanometer of period 0·2 sec. The galvanometer was shunted with a resistance of 90 ohms to damp irregular oscillations of the mirror. The motion of the spot of light reflected from the galvanometer mirror was recorded photographically on a sheet of Insurance Bromide, Grade A, wrapped round a cylindrical drum driven by a small synchronous motor, the drum moving laterally as it rotated on its axis. This camera was designed and constructed by Mr J. A. Linton, of the workshop staff. After considerable experience had been gained in identifying very small kicks due to bursts, the drum was finally geared down to give a film speed of 7 mm./min., and a single sheet of bromide paper then sufficed for a 24 hr. run. A high-resistance leak was connected between the control grid and the filament of the valve, so that after the sudden kick produced by a burst, the trace on the film returned to the zero line with a half-time of about 8 sec. The kicks on the film were subsequently measured with a travelling microscope carrying a vernier reading to 0·01 mm., and kicks smaller than 0·40 mm. were rejected as falling below the limit of certain detection. While the counts may have included small spurious kicks due to statistical fluctuations or to groups of contamination alpha-particles, it will be clear from the nature of the results obtained that any such kicks formed a relatively unimportant background of constant magnitude, on which were superposed larger real effects of particular interest.

In order to estimate the sensitivity of the apparatus to shower particles, certain constants of the apparatus had to be determined. The capacities of the large and of the small chamber, together with that of the shielded lead to the grid of the electrometer valve, were found to be 27 and 15 cm. respectively. The mean path length of a cosmic-ray particle through each chamber was calculated with the aid of a simple mechanical model: this model permitted the direct measurement of the path length for a large number of regularly spaced angles of incidence and points of entry of the ray into the chamber, and each value was suitably weighted according to geometrical considerations and to allow for the drop in number of the shower particles with increasing angle of divergence from the vertical (Froman & Stearns 1938). The mean path length was found to be 25·2 and 9·7 cm. for the large and small chambers respectively when placed with their axes of symmetry vertical, and 17·0 cm. for the large chamber when placed with its axis horizontal. From these data a fairly accurate comparison of the sensitivities of the two chambers could be made.

Only the order of magnitude of the absolute sensitivities of the chambers could, however, be determined. In the first place the value of the mean specific ionization by cosmic-ray particles is somewhat uncertain: a mean value of 60 ion-pairs/cm. path in air at N.T.P. was adopted. In the second place it was not found easy to estimate the factor introduced by the amplifying system. The method first employed was to find the deflexion produced on the film by applying a given steady potential to the control grid of the electrometer valve (Mohr & Stafford 1942), but this method was later found to be unsatisfactory, probably because it does not closely reproduce the experimental conditions which exist when the amplifier is in operation with the ionization chamber. A more reliable method was adopted, which consisted in finding the deflexion produced on the film—actually 13 mm.—when a small potential was induced on the collecting electrode of the small chamber as a result of suddenly altering the potential applied to the chamber electrodes by 0·1 V. Account was taken of the fact that the capacity of the shielded lead from the chamber to the valve was approximately equal to the capacity of the small chamber, and that on the average about one-fifth of the total charge collected by the chamber electrode has leaked away through the grid leak by the time the collection of the ions is complete. It was then calculated that a 1 mm. kick corresponds approximately to a burst of 90 particles traversing the large chamber, or 10 particles in the small chamber, when the axis of each chamber is vertical.

When investigating bursts under thick layers of iron or lead, a square thick-walled enclosure of the same material was built up round the sides of the chamber to prevent electron showers from the walls and ceiling of the room reaching the chamber. The layer of material directly above the chamber overlapped the walls of the enclosure by a few cm. Since the precise geometrical arrangement was found to be of some importance, that used in most of the runs with iron is illustrated in figure 1. The arrangement of the lead was similar, except that the thickness of the enclosure walls was 10 cm. instead of 14 cm. as for iron.

RESULTS

The results for the rate of occurrence of bursts of different sizes under different conditions are given in table 1. Their most interesting feature is found to lie in the manner in which the rate varies with the thickness of iron or lead above the chamber. This variation is shown in figures 2 and 3.

TABLE 1. TOTAL NUMBER OF BURSTS CORRESPONDING TO KICKS OF DIFFERENT SIZES

L or S denotes whether the large or small chamber was used;
V or H whether it was placed with its axis vertical or horizontal.

Type of chamber ...	L	L	L	L	L	L	L	L	L	L	S	L	L	L	L
Axis of chamber ...	V	V	V	V	V	V	V	V	V	V	V	V	H	H	H
Shielding material ...	Pb	Pb	Pb	Pb	Fe	Fe	Fe	Fe	Fe	Fe	Fe	Fe	Fe	Fe	Fe
Cross-section of enclosure (cm.^{2}) ...	730	730	730	730	730	730	730	730	730	730	400	3500	2020	2020	2020
Thickness of material above chamber (cm.)...	10	20	30	40	14	30	46	60	76	90	46	14	14	30	46
Time (hr.) ...	234	135	142	113	131	134	134	99	106	78	507	103	47	74	59
No. of kicks between															
0·4– 0·5 mm.	114	100	93	55	93	99	86	97	122	85	116	51	102	101	72
0·5– 0·6 mm.	73	74	66	38	41	55	44	57	72	68	90	30	46	34	23
0·6– 0·7 mm.	46	35	36	28	25	21	34	24	55	18	53	5	21	10	16
0·7– 0·8 mm.	29	21	26	18	20	18	24	18	31	14	43	6	16	9	6
0·8– 1·0 mm.	44	34	33	22	15	24	28	23	29	13	37	6	15	6	3
1·0– 1·2 mm.	20	16	14	15	15	11	19	7	9	13	37	—	8	5	3
1·2– 1·4 mm.	16	12	15	8	10	10	6	5	5	9	28	1	3	—	2
1·4– 1·8 mm.	21	11	9	11	6	10	10	11	11	3	19	—	1	3	1
1·8– 2·2 mm.	6	9	6	3	4	3	7	7	3	4	9	—	1	1	1
2·2– 2·6 mm.	9	2	7	3	2	5	—	3	5	2	7	—	—	1	—
2·6– 3·0 mm.	4	4	5	2	3	5	3	1	3	1	4	—	1	—	1
3·0– 4·0 mm.	4	7	3	4	3	1	—	1	3	3	6	—	2	1	1
4·0– 5·0 mm.	3	1	1	2	—	—	1	—	2	2	2	—	—	—	1
5·0– 6·0 mm.	2	2	1	1	—	—	—	1	1	—	1	—	—	—	—
6·0– 8·0 mm.	2	1	1	1	2	—	—	1	—	—	2	—	2	1	—
8·0–10·0 mm.	—	1	—	—	—	—	—	—	—	—	1	—	—	—	—
Over 10·0 mm.	2	—	1	—	1	2	2	—	—	—	1	—	1	—	1

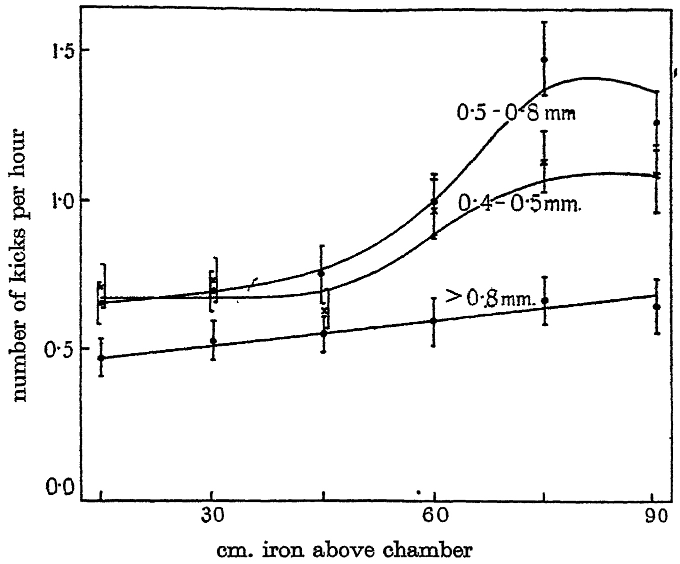


FIGURE 2. Rate of occurrence of bursts in large chamber under iron. (The limits of error refer to the standard deviation.)

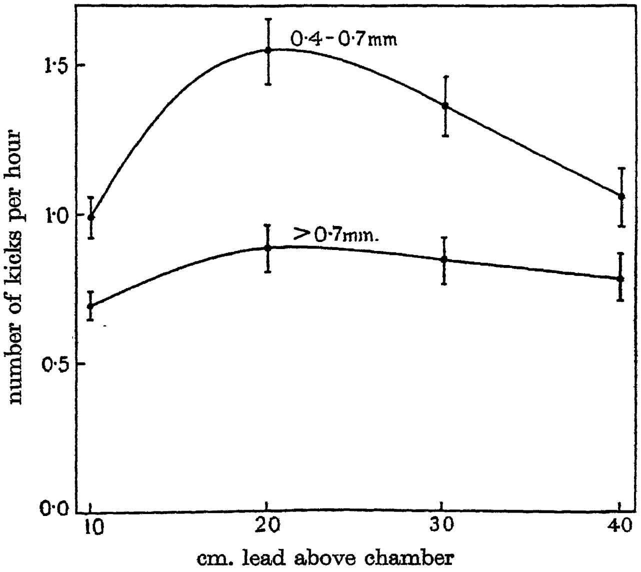


FIGURE 3. Rate of occurrence of bursts in large chamber under lead. (The limits of error refer to the standard deviation.)

It will be seen that pronounced maxima occur for the smallest bursts at about 20 cm. for lead and 75 cm. for iron, as against the values 17 and 30 cm. respectively at which the second maximum is found to occur in the Rossi curves obtained with counters. The statistical error of the present results is not small enough to make it certain that there is not a slight maximum at 30 cm. in our curves for iron in addition to the maximum at 75 cm. In this connexion it would have been of interest to have extended the curve for lead beyond 40 cm. to see whether there was not a further maximum, and it is unfortunate that it was not possible to do this with the amount of lead readily available to us.

While the maxima in our curves are presumably of the same nature as those found for two-particle showers by Swann & Ramsay (1940*a*) and by Broussard & Graves (1941), our curves for kicks of 0·4–0·5 mm. correspond on the average to the passage of about 40 particles through the chamber.

Light was thrown on this point by taking runs with the smaller chamber under 46 cm. of iron and comparing the burst rates for the large and the small chamber. The comparison is best effected by plotting, as is usual in this work, the logarithm of the total number of kicks greater than a certain size against the logarithm of the size of the kick (figure 4). The data show that a kick of 0·4 mm. with the large chamber corresponds on the average to a burst comprising the same number of particles crossing the chamber as a kick of 3·5 mm. with the smaller chamber. We then find from figure 4 that kicks of over 0·4 mm. with the large chamber are about 100 times as frequent as kicks of over 3·5 mm. with the small chamber, whereas the ratio of the cross-sections of the chambers is only about 5. Hence the usual assumption that the burst rate under large thicknesses is proportional to the cross-section of the chamber, which is based in turn on the assumption that the bursts are due to fairly narrow showers diverging from a comparatively small area in the material above the chamber, is certainly not justified. The greatest caution is therefore necessary in comparing experimental data on the absolute frequency of large bursts with theoretical values, in attempts to determine the spin or other properties of the mesotron (Christy & Kusaka 1941).

Turning again to figure 4, it will be seen that while most of the points lie approximately on a straight line for each chamber, deviation from linearity occurs for the smallest bursts, as shown by the dotted lines. Such deviations are to be expected in view of the anomalous variation of burst rate with thickness for the smallest bursts. Furthermore, it is reasonable to expect that the point at which the deviation begins should in each case correspond to a shower comprising, on the average, the same total number of particles in all, N (say), but that the proportion of the rays entering the chamber will depend on the size of the latter, if the shower extends over a region comparable with the cross-section of the chamber. The bend in the curve for the large chamber occurs for a kick of 0·7 mm., corresponding to the passage of about 60 rays through the chamber, while for the small chamber the bend occurs for a kick of 1·0 mm. corresponding to the passage of about 10 particles through the chamber. This indicates that when an N-particle shower takes place above either chamber, approximately six times as many of the shower particles pass through the large chamber as through the small, and since the ratio of the cross-sections of the space enclosed between the electrodes is 5·4, it suggests that the shower particles extend over an area at least as large as that of the circular section of the large chamber.

Considering now the observed frequency of occurrence of bursts due to these N-particle showers, figure 4 shows this to be 0·80 per hour for the large chamber and 0·22 per hour for the small chamber, or a ratio of 3·6. This value lies between the ratio, 5·4, of the cross-sections of the space between the electrodes for each chamber, and the ratio, 1·8, of the areas enclosed by the iron surrounding the chambers in each case, which is a reasonable result.

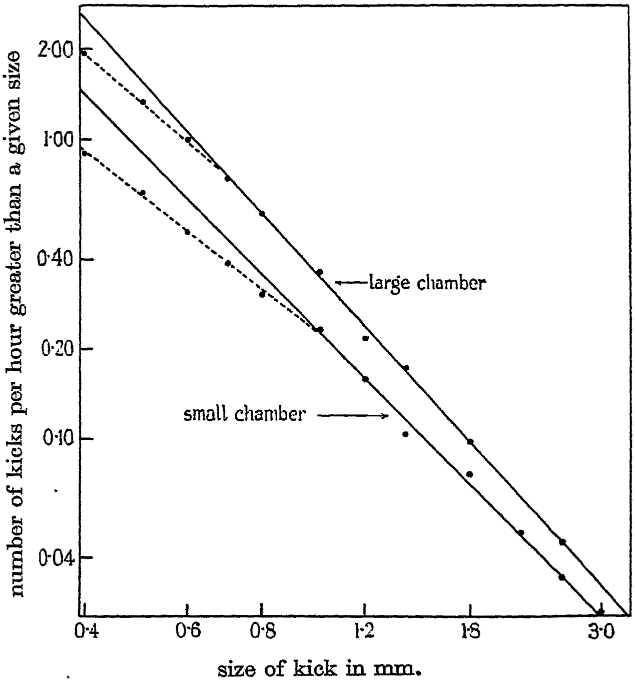


Figure 4. Burst production in chambers of different sizes under 46 cm. iron. Sensitivity of large chamber, 90 particles per mm. kick; of small chamber, 10 particles per mm. kick.

The results, then, can be considered as due to burst particles originating, not just from one small region in the material above the chamber, but from different regions in this material and probably also in the material round the sides of the chamber. The part played by the walls was investigated by increasing the horizontal cross-section of the iron enclosure round the large chamber to five times the value previously used, and at the same time increasing the area of the layer of iron above the chamber until it overlapped the walls of the enclosure, while keeping the thickness of the enclosure walls and the height of the material above the chamber the same. It was found that the frequency of occurrence of the smallest bursts was reduced to about one-half, while that of the larger bursts was reduced to a much smaller fraction, only one of the 99 kicks recorded being over 1 mm. in size.

It is obvious, therefore, that the walls of the smaller enclosure scatter many of the shower particles back into the chamber, the more so since the increased burst rate due to the greater proximity of the walls more than counterbalances the decrease in frequency to be expected from the smaller area of burst-producing material above the chamber. As for the largest bursts, their number is so greatly increased by the proximity of the walls, that the particles incident on the enclosure walls must surely be producing a number of secondary particles in the latter. The process may be regarded as one of the whole chain of events starting with a single mesotron and ending with burst particles, so that the material of the enclosure walls—just as much as the material above the chamber—may be considered as a region from which some of the groups of shower particles originate. Further evidence for the effect of the size of the enclosure on the burst rate is provided by a comparison of the burst rates with the large chamber and with the small chamber under 46 cm. of iron. The somewhat smaller enclosure used with the small chamber is seen from figure 4 to result in a less rapid decrease in the burst rate with increasing size of burst than with the large chamber.

The geometrical arrangement of the large chamber with its surrounding material was finally altered by placing the chamber with its axis horizontal, surrounding the sides and ends with the smallest possible rectangular enclosure of the same wall thickness, and placing above it thicknesses of 14, 30, and 46 cm. of iron in turn. Again a much more rapid decrease in the frequency of occurrence of the larger bursts with increasing size of burst was observed than in the case of the upright chamber with its smaller enclosure. It was also seen that the burst frequency—thickness curves were similar in form between 30 and 46 cm. (though relatively higher at 14 cm.) to those obtained for kicks of the same size when the chamber was erect. This indicates that although the mean path length of particles passing through the chamber is reduced by a factor of 0·7 when the latter is placed on its side, approximately the same total ionization is observed with either position of the chamber for showers comprising the same total number of particles, as is to be expected if the shower particles extend over a relatively large area.

Discussion

The appearance of a maximum in our curves at large thicknesses is a phenomenon which must be closely related to that of the second maximum observed by Swann and Ramsay (1940*a*) using trays of counters for two-particle showers in tin, by Broussard & Graves (1941) using a cloud chamber for two-particle showers in lead, and by many others with coincidence counters. There seems little doubt now that these results are all due to multiple processes involving the production of mesotron showers.

Heitler & Peng (1942) have shown that the probabilities of the production of showers of two or more mesotrons by a primary mesotron reach a maximum for certain definite values of the energy of the primary mesotrons; these values of the energy required to produce a maximum for showers of n particles increasing with n. The existence of mesotron showers has now been demonstrated by several experiments. Schein, Jesse & Wollan (1941), and Carlson & Schein (1941) have put forward evidence for the creation of mesotrons in multiples in the upper atmosphere. Swann & Ramsay (1940*b*), Damy de Souza Santos, Pompeia & Wataghin (1941) and Janossy (1942) have demonstrated with counters the production of mesotron showers in certain elements. Powell (1941) and Wollan (1941) have obtained cloud-chamber photographs showing the production of showers of several mesotrons in lead.*

The mesotrons composing such showers will give rise to fast electrons at some point in their path, and these fast electrons will in turn produce by the usual cascade process the electron showers which are observed as bursts under large thicknesses. Since the mesotron shower particles may diverge widely, they may give rise to electron showers in widely separated regions above and around the sides of the ionization chamber. This mechanism would account for the extensive nature of at least some—especially the large—electron showers observed with the large chamber, while at the same time giving the peaks in the burst frequency—thickness curves as characteristic of showers composed originally of a few mesotrons.

We have seen that the observed rate of production of bursts of different sizes depends very markedly on the geometrical arrangement of the chamber and the material surrounding it. The same geometrical dependence characterizes the results of some of the observers who have investigated the second maximum of the Rossi curve with counters (e.g. Schmeiser & Bothe 1938). To what extent the prominence of the peaks in the burst frequency—thickness curves depends on the particular geometrical arrangement is still to be investigated. Further experiments on this and other points are being carried out with the small chamber by one of us (C. B. O. M.).

In conclusion we wish to thank Mr J. A. Linton and Mr G. Laing of the workshop staff for setting up the pressure chambers and for much valuable assistance. We are also indebted to the University of Cape Town for a generous research grant for the purchase of apparatus, without which this research could not have been undertaken.

---

* It has been suggested that these showers may be of the explosive type, but there is no conclusive evidence for this.

REFERENCES

Bhabha 1938 *Proc. Roy. Soc.* A, **164**, 257.
Broussard & Graves 1941 *Phys. Rev.* **60**, 413.
Carlson & Schein 1941 *Phys. Rev.* **59**, 840.
Carmichael 1936 *Proc. Roy. Soc.* A, **154**, 223.
Christy & Kusaka 1941 *Phys. Rev.* **59**, 414.
Damy de Souza Santos, Pompeia & Wataghin 1941 *Phys. Rev.* **59**, 902.
du Bridge 1933 *Rev. Sci. Instrum.* **4**, 532.
Froman & Stearns 1938 *Rev. Mod. Phys.* **10**, 133.
Heitler 1938 *Proc. Roy. Soc.* A, **166**, 529.
Heitler 1941 *Proc. Camb. Phil. Soc.* **37**, 291.
Heitler & Peng 1942 *Proc. Camb. Phil. Soc.* **38**, 296.
Janossy 1942 *Proc. Roy. Soc.* A, **179**, 361.
Mohr & Stafford 1942 *Nature, Lond.*, **149**, 385.
Nereson 1942 *Phys. Rev.* **61**, 111.
Powell 1941 *Phys. Rev.* **60**, 413.
Schein, Jesse & Wollan 1941 *Phys. Rev.* **59**, 615.
Schmeiser & Bothe 1938 *Ann. Phys., Lpz.*, **32**, 161.
Swann & Ramsay 1940*a* *Phys. Rev.* **58**, 477.
Swann & Ramsay 1940*b* *Phys. Rev.* **57**, 1051.
Wilson 1941 *Proc. Camb. Phil. Soc.* **37**, 301.
Wollan 1941 *Phys. Rev.* **60**, 532.

Applications of the photographic method to problems in nuclear physics

II. The scattering of 8·8 and 13 MeV neutrons by protons

By F. C. CHAMPION, *King's College, University of London*, AND
C. F. POWELL, *The H. H. Wills Physics Laboratory, University of Bristol*

(*Communicated by A. M. Tyndall, F.R.S.—Received 22 September* 1943)

A method is described for determining the orientation of proton tracks in a photographic emulsion with an accuracy of ± 1°. This technical development increases the range of application of the photographic method to problems in nuclear physics. The method has been applied to an investigation of the scattering of 8·8 and 13·0 MeV neutrons by protons.

If $I(\phi)$ is the intensity per unit solid angle of the scattered neutrons in the centre of mass co-ordinate system of the colliding particles, the measurements give values of $I(\phi)$ in the range from $\phi = \pi$ to $\phi = \frac{1}{2}\pi$. When this range in the value of ϕ is divided into five equal intervals, the observed numbers of tracks in the different intervals are consistent with the scattering being isotropic in the centre of mass system to within the limits of the statistical error, for the 8·8 MeV group of neutrons. There is some evidence for a periodic fluctuation of high order in the angular distribution of the scattered intensity which remains to be confirmed. For the 13 MeV group the ratio of $I(\pi)/I(\frac{1}{2}\pi)$ is found to be $0{\cdot}67 \pm 0{\cdot}20$.

The neutrons employed in this experiment were produced by bombarding boron with 600 keV deuterons and arise from the reaction $B^{11}(d, n)\,C^{12}$. There is evidence that the excited state of C^{12} at 4·5 MeV is complex.

1. INTRODUCTION

The two principal objects of the present paper are, first, to give an account of the methods we have developed in the course of the past three or four years to determine accurately the spatial orientation of the track of a charged particle in a photographic emulsion; and, second, to describe the application of these methods to the investigation of the scattering of fast neutrons by protons. Preliminary reports on the results have already been published (Powell 1940; Powell, Heitler & Champion 1940).

Hitherto experiments with the photographic method have been for the most part confined to the measurement of the range of protons and α-particles, and for this purpose the tracks are arranged to be inclined at only small angles with the plane defined by the surface of the emulsion. The importance of the first of our objectives lies in the fact that if the angle of dip of a track into the emulsion can be determined with precision, the analogy of the method with the expansion chamber becomes more complete, and the scope of the method for experiments in nuclear physics is increased. In this paper we show that the angle of dip of a particle into the emulsion can be determined with an accuracy of about $\pm 1^\circ$.

The importance of the study of neutron-proton scattering at high energies lies in the fact that it is highly desirable, for the development of nuclear physics, to determine the character and the magnitude of the forces between the neutron and the proton at distances corresponding to the very close approach of the two particles. Such a determination can be most easily made at present by a study of neutron-proton scattering at high energies. Various theoretical papers (Massey & Buckingham 1937; Rarita & Schwinger 1941) indicate that, with increasing neutron energy, anomalies in the scattering may be expected to appear rather suddenly in the region of 15 MeV.

2. PREVIOUS EXPERIMENTS

Experiments on neutron-proton scattering have been made by a number of investigators using the homogeneous group of neutrons which is emitted when deuterium is bombarded by deuterons with an energy of the order of 600 keV. The energy of the neutrons from such a source is of the order of 2·5 MeV. The best experiments, for example those of Dee & Gilbert (1937), indicate that with neutrons of this energy the scattering is isotropic in the centre of mass co-ordinate system of the colliding particles. To avoid circumlocution we refer to this as isotropic scattering.

Since the preliminary report on the present work was published, Tatel (1942) has given an account of experiments on the same subject using a neutron source produced by the bombardment of beryllium by high-energy deuterons from the cyclotron. As with many of the light elements, beryllium, when bombarded by deuterons with an energy of ~ 8 MeV, emits neutrons with a continuous distribution in energy. Tatel arranged to observe the recoil protons produced by the neutrons in the high-energy tail of this continuous distribution, detecting them with an ionization chamber and linear amplifier. He made observations at two angles only, 16 and 45°, to the direction of the primary neutrons, and concluded that the results were in accord with the assumption that at 11 MeV the scattering is isotropic.

Finally, in a recent theoretical paper by Hulthen (1943) on this subject, there is a reference to experiments by Amaldi, Bocciarelli, Feratti & Trabacchi (1942). Let $I(\pi)$ and $I(\frac{1}{2}\pi)$ be the intensities of the scattered neutrons in the centre of mass system of co-ordinates of the colliding particles at 180 and 90° respectively. Amaldi finds the ratio $I(\pi)/I(\frac{1}{2}\pi)$ to be between the values 0·71 and 0·52 for neutrons of average energy between 12·5 and 14·0 MeV. These values are taken from Hulthen's paper, as we have not access to Amaldi's publication.

3. THE NEUTRON SOURCE

It is evident that the most desirable source for experiments of this type is one in which the neutrons are homogeneous in energy as is the case with the D + D neutrons at a given angle of emission to the direction of the primary beam. No such source is obtainable, however, at the present time for the range of energies from 8 to 15 MeV. The best conditions of measurement available in practice are provided by a source emitting a number of homogeneous groups of neutrons. If the difference in energy between the different groups is sufficiently great we can ensure that, over a certain range in the angle of scattering, the proton recoils produced by neutrons in the different groups can be distinguished because of the difference either in their range or in the total ionization which they produce. The distribution in angle of the intensity of the projected protons arising from the different groups can then be determined in this range of angles.

Two neutron sources of this character are well known; those produced by the bombardment of boron or of fluorine by deuterons with an energy of from 500 to 1000 keV, the deuterons being accelerated by d.c. high-tension generators. In the present work we have preferred to use the neutrons from boron because the high-energy neutron groups are more widely separated in energy than is the case with the fluorine target. The boron target emits, amongst others, two groups of fast neutrons of mean energy 8·8 and 13·0 MeV in a direction at right angles to the incident deuterons, the spread in energy in each group being less than 1·5 MeV. For the present work we have used the plates obtained in the experiments previously described (Powell & Fertel 1939; Powell 1943) in which a boron target was bombarded by the 600 keV deuteron beam delivered by the Bristol generator.

Two advantages, for the present experiments, of using a neutron source based on a high-tension generator may be emphasized. First, with proper adjustment of the voltage distribution across the electrodes of the main accelerating tube, and with careful alinement of these electrodes, the deuteron beam can be sharply focused on to the target. In the present experiments the area of the target from which the majority of the neutrons originate is a circular patch with a diameter of about 4 mm.

It is therefore possible to secure conditions of good geometry without long exposures and the attendant dangers of a general background fog due to the X-rays, γ-rays and neutrons coming from various parts of the apparatus.

Secondly, with relatively slow deuteron bombardment, the number of neutrons in the groups of highest energy, corresponding to the formation of the product nuclei in the ground or lower excited states, is of the same order of magnitude as that in the slower neutron groups which are present. This has the consequence that the long recoil tracks in which we are interested do not have to be distinguished against an excessively large background of unwanted shorter tracks produced by slower neutrons. In contrast with the favourable conditions which are thus provided, plates obtained with a source produced by the bombardment of the light elements by the high-energy deuteron beams delivered by the cyclotron show very few tracks due to neutrons of the highest energy. It will be shown in a subsequent paper that the neutrons from such sources show a continuous distribution in energy with a maximum intensity in the energy region between 3 and 5 MeV and a very rapid fall in going to higher energies. The high-energy neutrons corresponding to the formation of product nuclei in the ground state are very rare. An additional disadvantage of such a source is that the plates after development show a general background of grains, due to the passage through the plate, during exposure, of slow neutrons and γ-rays from the cyclotron tank, and this background contributes to the difficulty of distinguishing the long tracks.

These considerations are presumably of some significance for work with counters, and they are decisive for experiments employing the photographic method. The smaller yield of neutrons from the high-tension generator as compared with the cyclotron is of no consequence, since even with the former the exposure times necessary are very short.

4. DETERMINATION OF THE SPATIAL ORIENTATION OF A TRACK IN THE EMULSION

The basic problem to be solved in determining the spatial orientation of a track in the emulsion is the measurement of the angle of dip of the track; i.e. the angle the track makes with the plane defined by the surface of the emulsion. Consideration of the restricted depth of focus of high-aperture objectives, $0{\cdot}3\,\mu$, suggested that it might be possible to make such determinations by measuring the depth below the surface of the emulsion of different grains in the track. Our microscopes are fitted with graduations on the fine adjustment wheel, but the makers informed us that this graduation does not provide an accurate method of making depth measurements owing to the design of the mechanical system on which the fine motion is based. Tests of this opinion using the best modern microscopes have confirmed it. We therefore fitted a precision depth gauge in the way illustrated in figure 1.

The gauge is fixed to the upper arm of the optical bench of the microscope which carries the objective and eyepieces. The pin of the gauge, P in figure 1, bears on a ground surface, S, of hardened steel which is attached to the mechanical stage, which in turn carries the photographic plate. The gauge therefore indicates any relative movement of the focal plane of the objective and the photographic plate.

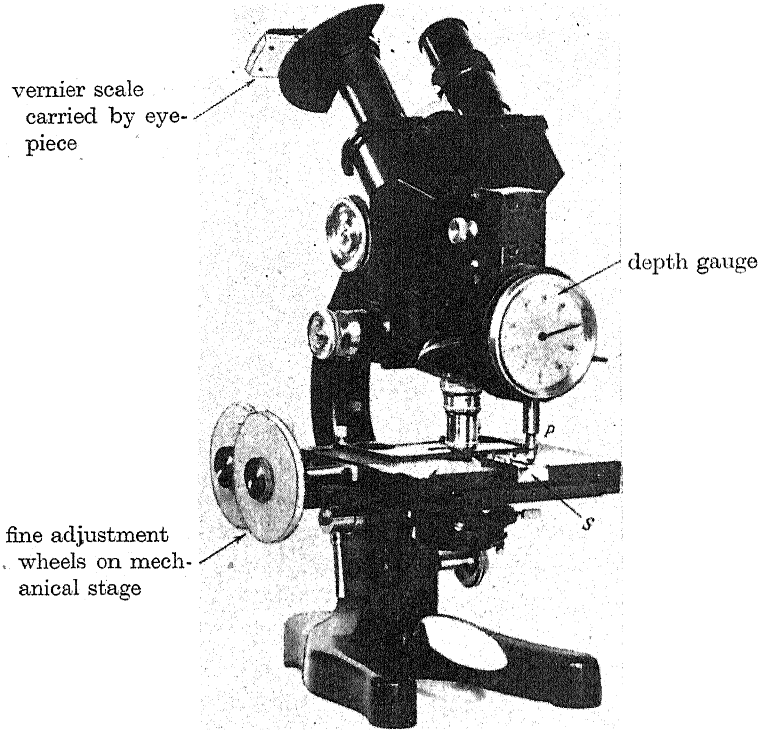

vernier scale carried by eyepiece
depth gauge
p
S
fine adjustment wheels on mechanical stage

FIGURE 1. Microscope fitted with a depth gauge and vernier scales.

The gauge is graduated in divisions representing 0·0001 in., and it was found possible to set a particular grain in focus with a consistency within 0·4 division, i.e. to within 0·001 mm., or 1μ, when using the optical equipment previously described (Powell 1943). The actual distance between two grains which appear, when viewed through our microscope, to be separated by the full length of the eyepiece scale, is 0·1 mm. One would therefore expect to be able to determine the angle at which a long track 'dips' with respect to the plane of the emulsion with an accuracy of the order of ± 1°. This estimate takes no account of the fact that even the fastest proton tracks suffer small-angle deflexions from their initial direction in passing through the emulsion, so that the actual accuracy achieved will be less than this value.

The second measurement necessary in order to define the direction of the track relative to the emulsion is the angle which the projection of the track on the plane of the plate makes with some standard direction. This is done by setting a line in the eyepiece scale to coincide in succession with the separate grains in the track as the microscope is focused on them in turn. The setting is made by rotating the eyepiece which carries the scale and the fiducial line. This eyepiece is provided with a vernier moving over a graduated scale, and by this means the orientation of the eyepiece can be read to 0·1° and the setting can be made with an accuracy of 0·25°.

For making these measurements with the highest accuracy, it is important that proper attention should be paid to the adjustments of the substage illumination. If the illumination is not properly centred, and if it is not adequate for the high aperture of the objective, the full resolving power is not achieved and the accuracy of the individual depth measurements is reduced.

A second point requiring consideration is the fact that the field of highly corrected apochromatic objectives is not flat. This has the consequence that when using the objectives in conjunction with low-power periplanatic eyepieces, such as those normally employed, the grains in a track running across the centre of the field and precisely parallel to the surface of the emulsion are not all in focus at the same time. In determining the angle of dip of a track from measurements on the depth of two chosen grains it must therefore be arranged that the two grains are equidistant from the centre of the field. This is a sufficient condition for a reliable observation, since the curvature of the field is axially symmetrical.

In some emulsions, using very small plates with an area of 2–3 cm.^{2}, we have found that over some areas of the plate the tracks all show a slow continuous change of direction with depth. This effect only occurs rarely, and we believe it is due to the fact that the emulsion before development is under strain introduced during the process of manufacture, as the plate is drying, and to the release of these strains during development when the plate is wet. It is therefore desirable to use larger plates and to confine measurements to the central areas. Distortion of the gelatine is then negligible.

5. DETERMINATION OF THE ENERGY OF THE PROJECTED PROTONS AND THE ANGLE OF SCATTERING

Let OP of figure 2 represent a track in the emulsion. OA, which lies in a plane $OABC$, parallel to the surface of the emulsion, is the direction of the neutron stream through the point O during the exposure. Q is the projection of P on the plane $OABC$, and R is the projection of Q on the line OA. In measuring a track, the fiducial line carried by the rotating eyepiece is first set so that as the successive grains in the track are brought into focus they are seen to coincide with the line. In practice the ends of the graduations of the eyepiece scale are used as a fiducial line, and the beginning of the track is set to coincide with the beginning of the scale. The angle through which the eyepiece has been rotated from the standard position in which it is parallel to OA can then be determined from the reading of the vernier. This angle, which we call the angle of inclination, α, is the angle ROQ in figure 2. The difference in depth of the first grain in the track at O and the grain at the point P is determined from the readings of the depth gauge and is the length PQ in figure 2. Let $PQ = d$ and let the length OQ, measured directly by the eyepiece scale, be l. The angle of dip of the track, β, is then equal to $\tan^{-1}(d/l)$.* It is required to determine the angle ROP which the track makes with the original direction of the neutron stream through O, together with the length L of the track. If $R\hat{O}P = \theta$, then

$$L = l \sec\beta, \tag{1}$$

$$\cos\theta = \cos\alpha \cos\beta. \tag{2}$$

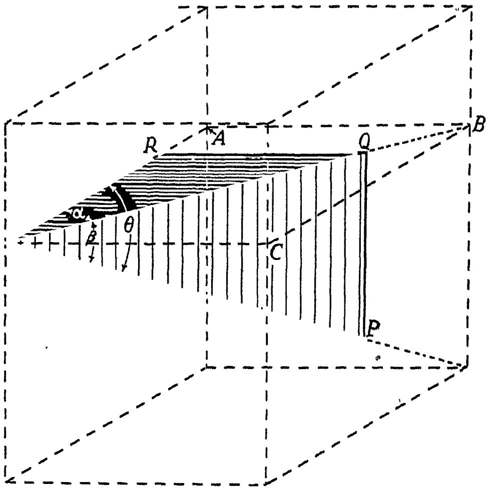


FIGURE 2. Definition of the angular characteristics of a track.

Knowing the stopping power of the emulsion and the range-energy relationship, the length L of the track enables one to determine the energy, E, of the proton producing it. The energy, E_0, of the neutron giving rise to the recoil proton is then given by the equation $E_0 = E\sec^2\theta$. It is thus possible to deduce the energy distribution of the original neutron beam from observations of all proton tracks and not merely from those projected forward within a small angle of the direction of the beam. The accuracy of the determination of E_0 must be expected to decrease rapidly with increasing values of θ according to the relation

$$\delta E_0 = 2E_0 \tan\theta\, \delta\theta.$$

Since it is required to measure the relative probability of scattering in different directions, it is necessary to ensure that all tracks within closely defined limits are measured, and checks against any tendency of an observer to choose tracks with particular characteristics must be applied. For this purpose maps are constructed of those areas of the plate on which measurements are made. In addition to the angle of inclination, these maps show the co-ordinates of the beginning and end of each track as deduced from the readings of the verniers on the x and y movements of the mechanical stage. For accurate mapping it is necessary to determine the x

---

* In practice the angle β is determined from the difference in depth of two grains, near the beginning of the track and separated by the length of the eye-piece scale (0·1 mm.).

and *y* co-ordinates of a point in the plate with greater precision than is allowed with scales and verniers normally supplied with the mechanical stage, which read to 0·1 mm. The milled knobs controlling the movements of stage were therefore fitted with graduated wheels with suitable vernier scales as shown in figure 1. The maps allow the observer to re-examine a given plate for any tracks that he may have overlooked and allow one to compare the efficiencies of different observers in this respect. In practice it is found that with experienced observers the number of tracks missed is very small, and the differences between the measurements of different observers on the same track are trivial.

6. ENERGY DISTRIBUTION OF THE NEUTRONS

An area of 200 mm.^{2} has been examined by the methods described in the previous paragraph, and 4000 tracks contained in it, which were produced by neutrons with an energy greater than 3·5 MeV, have been mapped. Tracks produced by neutrons of lower energy were ignored. This total area was made up of two nearly equal parts taken from two separate plates which were similarly disposed during exposure with respect to the axis of the primary deuteron beam. The two sets of measurements, *A* and *B*, were treated separately in the following way. In set *A*, the calculations were carried through for all tracks produced by neutrons with an original energy greater than 3·5 MeV. In set *B*, the calculations were restricted to those tracks which correspond to an original neutron energy greater than 7 MeV. From the measurements on the tracks from the two sets which have values of θ less than 10°, together with the measurements on similar plates taken with the boron target described in the previous paper, it is possible to make an estimate of the homogeneity of the neutrons in the different groups forming the neutron spectrum emitted from the target, and the results are shown in figure 3.

It will be seen from this figure that the width at half-maximum of the 13 MeV group, which corresponds to the formation of C^{12} in the ground state, is 0·3 MeV, a value which is in good agreement with the results obtained in a previous paper (Powell 1943), where it was shown that the experimental width at half-maximum of homogeneous groups of protons was nearly constant over the energy range from 3 to 10 MeV and equal to 0·25 MeV.

In contrast with this result the width of the 8·8 MeV group is 0·8 MeV, a value more than three times as great as that obtained with the protons of nearly the same energy, coming from the reaction $B^{10}(d,p)B^{11}$ (Powell 1943). It is therefore clear that the neutrons in this group are not homogeneous in energy. This spread in energy cannot be due to a short lifetime in the excited state of C^{12} at 4·6 MeV, since this state is not of sufficient energy to lead to the disintegration of C^{12} into three α-particles. The most reasonable explanation would therefore seem to be that the state is a multiplet. The form of the neutron distribution is compatible with the state being a doublet with a separation of about 0·5 MeV. The mean energy of the excited state as deduced from the present experiments, 4·5 MeV, is rather higher than that obtained by other observers. Thus Cockcroft & Lewis (1936) obtained the value 4·2 MeV from a study of α-particles emitted from the reaction $N^{14}(d,\alpha)\,C^{12}$, whilst Bonner & Brubaker (1936) give the value 4·4 from an examination of the neutrons from the reaction used in the present work. The Q value obtained in the present work for the production of C^{12} in the ground state, 13·6, is in good agreement with that expected from the known masses of the nuclei involved, viz. 13·68. For the purposes of the present experiments it is sufficient to note that the width of peaks at half-maximum is less than 0·9 MeV, and the full width less than 1·5 MeV.

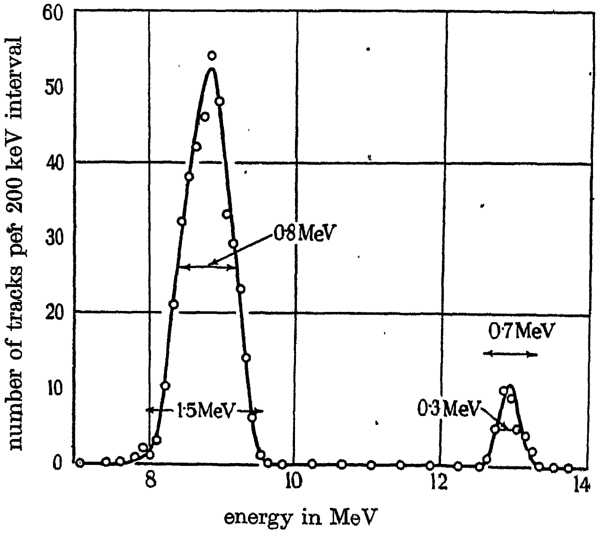


FIGURE 3. Distribution in energy of the groups of fast neutrons.

7. ACCURACY OF THE ANGULAR MEASUREMENTS

It has already been seen that the estimated errors in the determination of the angle of inclination of a straight track is of the order of 0·25° and in the angle of dip, 1°. It is shown later that it is desirable to employ only those tracks which make angles of dip less than 10°. Since the actual angle of scattering, θ, is given by the relation $\cos\theta = \cos\alpha\cos\beta$, it follows that for most of the tracks the greater part of the error in the determination of θ will come from errors in the determination of α, so that the errors in the determination of the orientation of a track can be expected to be less than ± 1°. Thus

$$\cos\theta = \cos\alpha\cos\beta,$$

and hence

$$\delta\theta = (\cos\alpha\sin\beta\,\delta\beta + \cos\beta\sin\alpha\,\delta\alpha)\operatorname{cosec}\theta.$$

Putting

$$\alpha = 20°, \quad \beta = 5°, \quad \delta\alpha = +0{\cdot}25° \quad \text{and} \quad \delta\beta = +1°,$$

then

$$\theta = 20°\,36' \quad \text{and} \quad \delta\theta = 0{\cdot}47°.$$

It is a satisfactory feature of the method that with the restriction $|\beta| < 10°$, the errors in θ due to errors in β are greatest where they are of least consequence. If the above value for $\delta\theta$ is substituted in the relation $\delta E_0/E_0 = 2\tan\theta\,\delta\theta$, then

$$\delta E_0/E_0 = 2(\tan\alpha\,\delta\alpha + \tan\beta\,\delta\beta).$$

Substituting the maximum errors $\delta\beta = \pm 1°$ and $\delta\alpha = \pm 0{\cdot}25°$, then it is found that the maximum value of $\delta E_0/E_0$ varies from 0·6 % at small angles to 1·4 % at 45°, errors much smaller than the spread in a peak due to straggling. Unfortunately, it is not possible to display the full potentialities of the method in the present experiments on account of the uncertainties in the precise direction of the neutron stream relative to the plates during the exposure, for these uncertainties lead to larger errors in the determination of θ than those involved in the determination of the precise orientation of a track in the emulsion.

The uncertainty in the direction of the neutron flux arises from two causes. First, the region of the target bombarded by the neutron stream is estimated to have had a diameter of 4 mm. Since the plate was at a mean distance of 62 mm. from the source there is an uncertainty in the direction of the neutron stream at any point in the plate of ± 2°. Secondly, no attempt was made during the exposure of the plates to define the orientation of the plate with respect to a line drawn from the centre of the plate to the centre of the source. The original exposures were made for the purpose of deducing the energy distribution in the neutrons from the source by the methods described in Part I, and for this purpose an error of 2° in the setting of the plates is of no consequence. We estimate the errors in the measurement of θ arising from these deficiencies in the experimental arrangement to be ± 3°.

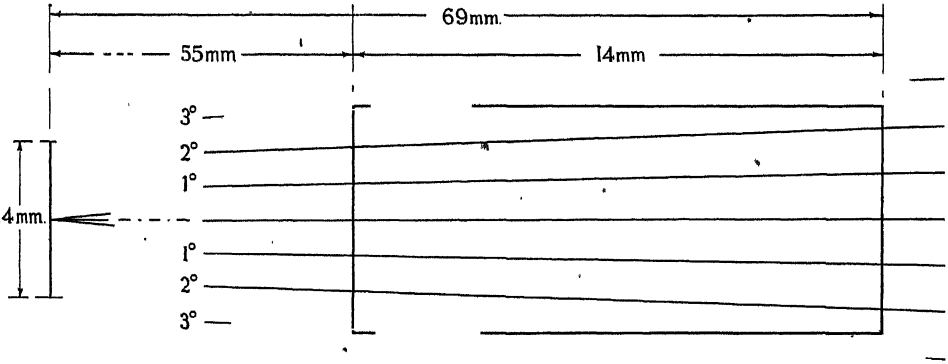


FIGURE 4. Variation in direction of primary neutron flux through the plate.

As a consequence of this source of error in the determination of θ we have not considered it worth while to make any allowance for the variation in the direction of the incident neutron stream in passing across the plate. The variation of this direction over the measured area is shown in figure 4, which is constructed on the assumption that the line joining the centre of the plate to the centre of the source runs parallel to the long edge of the plate. It remains to show that the uncertainties in the determination of θ are of no serious importance for the present experiments.

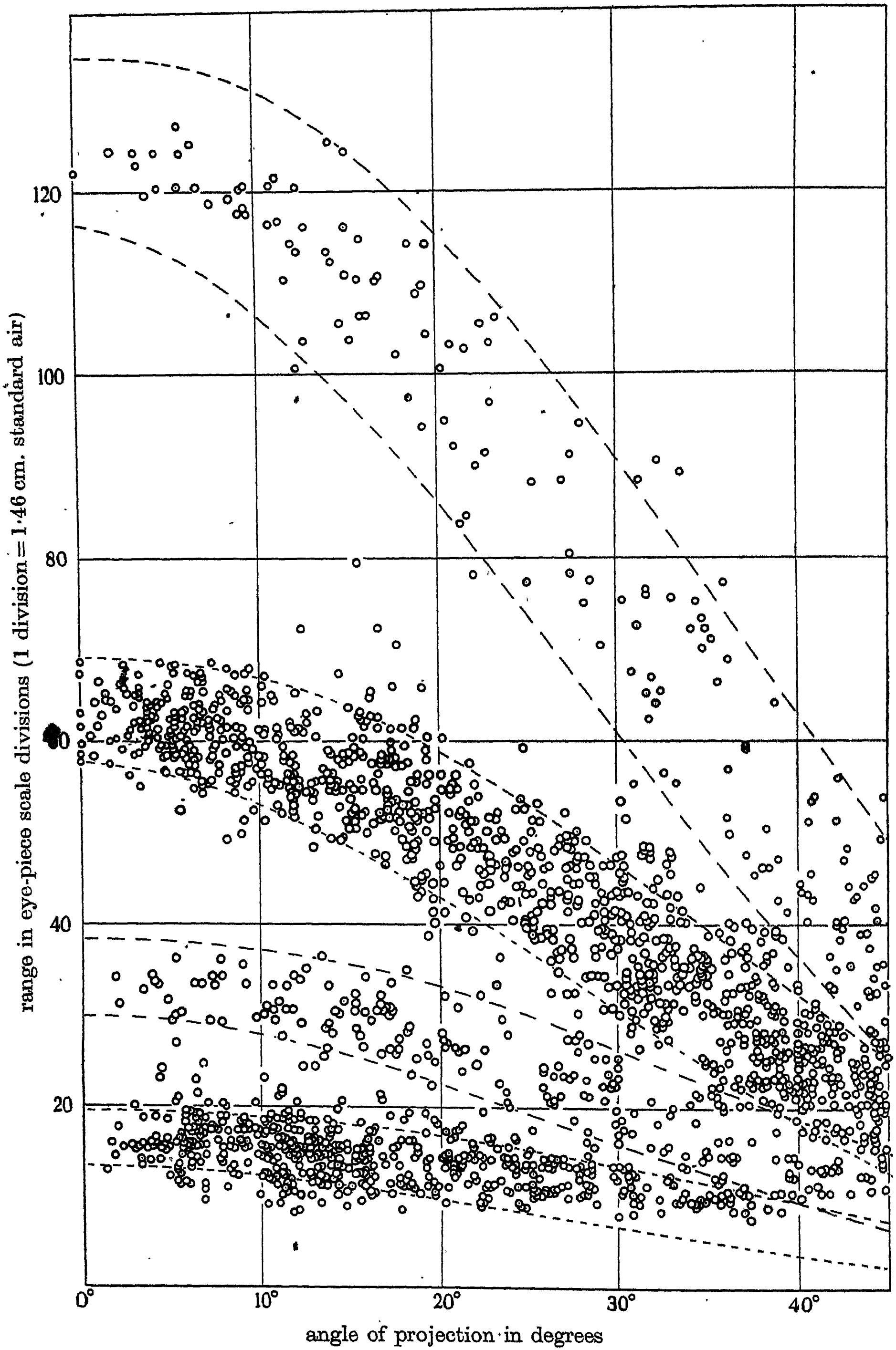


Figure 5. Distribution in range of protons with angles of projection between 0 and 45° to direction of primary neutrons.

8. TESTS OF THE ACCURACY OF THE MEASUREMENTS

In order to display the experimental results and to show that the accuracy in the determination of θ is within ± 3°, we show, in figure 5, the range of all measured tracks, of length greater than 12 cm. air equivalent, plotted against the corresponding angle of scattering, θ.

If there were no straggling of the proton tracks in the emulsion, and if the determinations of θ were made with perfect precision, the points representing the results, when plotted in the manner shown in figure 5, would be expected to lie on one of four lines, provided the four neutron groups were homogeneous in energy. The line for each group would be defined, (*a*) by the fact that the energy E of a proton, projected at an angle θ, is related to the energy, E_0, of the neutron producing it by the equation $E = E_0 \cos^2 \theta$ and, (*b*) by the range-energy relation for the emulsion. In figure 6, AB represents a portion of such a line. Owing to straggling, and to the inhomogeneity in the different neutron groups, it must be expected that in practice the lines will be broadened into bands with a total width corresponding at an angle θ, to a variation of ± 0·5 MeV in the proton energy. In figure 6, a is a point on AB corresponding to a particular value of θ, and b and c are plotted for proton energies 0·5 above and 0·5 MeV below the value corresponding to the point a. If the errors due to straggling were alone present, then most of the experimental points would be expected to lie in a band between two curves defined as the loci of b and c as a moves along AB.

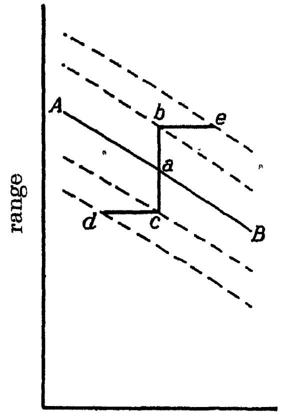


FIGURE 6. Effect of straggling and errors in the angular determinations on the range distribution at different angles of scattering.

A further spread in the experimental observations must be expected as a consequence of the uncertainty in the determination of θ. To take account of this the points d and e are plotted in figure 6 so that the lines dc and be have lengths corresponding to 3°. If the errors in the determination of θ are not greater than ± 3°, then the experimental results, corresponding to any one of the neutron groups, would be expected to lie within a band defined by the locus of d and e as a moves along the curve AB.

In figure 5 we include dotted lines giving the extent of the bands corresponding to each neutron group as defined in this way. It is clear that the great majority of the points lie within the bands, and that the bands corresponding to the three groups of highest energy do not seriously overlap, in the range from 0 to 45°. There is slight overlapping of the bands due to the 6 and 9 MeV groups between 40 and 45°, but the number of tracks common to both bands is small. It may therefore be concluded that in general the tracks which have been produced by neutrons from the 9 and from the 13 MeV groups can be distinguished without ambiguity.

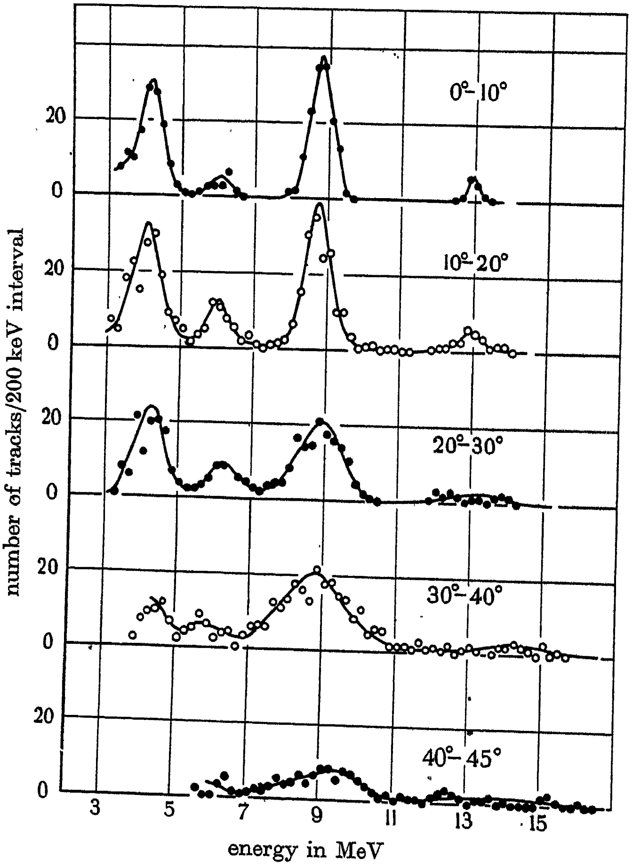


FIGURE 7. Distribution in energy of the primary neutrons as deduced from measurements on tracks in different intervals in the angle of scattering.

The experimental results can also be represented by plotting the distribution of the values of E_0 as calculated from tracks for which the value of θ lies in a given range of angles. Figure 7 shows such distributions in the energy as deduced from the measurements of all tracks in the ranges of angles 0–10°, 10–20°, 20–30°, 30–40°

and 40–45°. In this case a steady increase in the errors in the determination of E_0 must be expected with increasing values of θ, according to the relation

$$\delta E_0 = 2E_0 \tan\theta\, \delta\theta,$$

and the increasing spread in energy distribution corresponding to a given group is in satisfactory agreement with that to be expected.

Beyond 45° the overlapping of the bands is too serious to allow precise conclusions to be drawn about the origin of the tracks. It seems reasonable to assume that it would be possible to reduce the errors in the values of θ to $\pm 1°$, if proper care were taken to define the position of the plates at a sufficient distance from the neutron source. Figure 8 shows the bands for the three groups of highest energy calculated for an error of $\pm 1°$ in the value of θ. It will be seen that the bands for the 9 and 13 MeV groups remain separated up to 60° in this case. Our original plates were exposed in the middle of 1939, and further exposures under more satisfactory geometrical conditions should have been made, but with the outbreak of war the high-tension set was temporarily dismantled. In the original exposures the plates were placed close to the source in order to reduce the ratio of the neutron flux through the plates due to neutrons scattered from the walls, etc., to that from the source. Later experience has shown that this precaution was unnecessary. It would be an easy matter to place the plates 20 cm. from the source without seriously increasing the necessary exposure time. With the experimental arrangement described in Part I, this would reduce the uncertainty in the value of θ, due to the finite size of the source, to $\pm \frac{1}{2}°$.

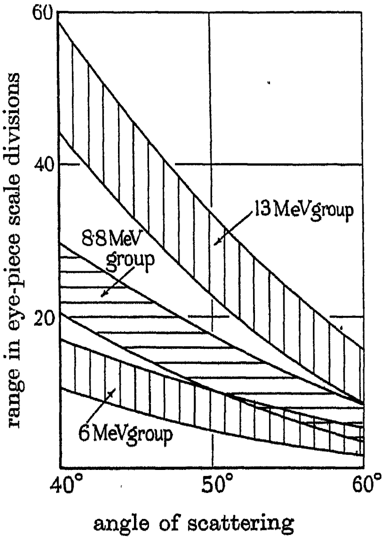


FIGURE 8. Improvement in resolution of range distribution of high-energy groups to be expected from improved geometry.

9. ESCAPE OF TRACKS FROM THE EMULSION AND CORRECTIONS FOR LOSS

In analysing the results a complication is introduced by the fact that the emulsion is of limited thickness, so that some tracks which start in the emulsion appear to end at the surface or at the glass. In general in such cases, the proton producing the track has left the emulsion before coming to rest, and therefore the length of the track does not give a measure of the proton energy. Such tracks are rejected since it is not possible to decide to which group the neutron responsible for the track belongs, and allowance must be made for this loss in calculating the scattered intensity at any angle from the observations. A method of treating this problem has been described by Richards (1941) for the case of an experiment in which the tracks accepted for measurement are those for which the angle of scattering is less than 12°.

Let D be the thickness of the emulsion. Ignore straggling and assume that a proton of energy E produces a track of length L. Let β be the angle of dip of a track of length L which starts in the surface of the emulsion, and which therefore ends at depth $L\sin\beta$, i.e. at a distance $D-L\sin\beta$ from the glass of the plate. It follows that of all the tracks produced by protons with characteristic E, β, which start in the emulsion, only those will end in the emulsion and be accepted for measurement which start within a depth $D-L\sin\beta$ from the surface of the emulsion. All other tracks of this class will pass out of the emulsion into the glass and will be neglected. It follows that only the fraction $f = 1-L\sin\beta/D$ of tracks produced by protons of this class will be measured. This calculation takes no account of the scattering of tracks through small angles, but the effect of scattering is small, since the probability of a proton being scattered so that it remains in the emulsion when, but for scattering, it would have passed out, is nearly exactly balanced by the reverse effect of particles being scattered out of the emulsion which would have otherwise remained in.

To examine the effects of the escape of protons from the emulsion we now represent the experimental results in a new way. We assemble together the measurements for all tracks which are identified as being produced by one of the two neutron groups of highest energy. As a simple criterion we attribute all tracks for which the calculated value of E_0 lies between 7 and 11 MeV to the 9 MeV group; and between 11 and 15 MeV to the 13 MeV group. We represent the results for all the tracks from one group in a diagram in which we plot for each track the quantity $\sin\beta$ as ordinate against $\cos\beta\sin\alpha$ as abscissa. In such a diagram reference to figure 2 shows that lines of constant angle of scattering, θ, are circles round the origin, and the figure is essentially a polar diagram representing the distribution in angle of the scattered particles. The results for the 9 MeV group are represented in figure 9.

If there were no loss of tracks at the surfaces of the emulsion and if there are no polarization effects associated with the neutron stream passing through the plates, one would expect the points representing the observations to be distributed with circular symmetry about the pole of the diagram, within the limits of statistical

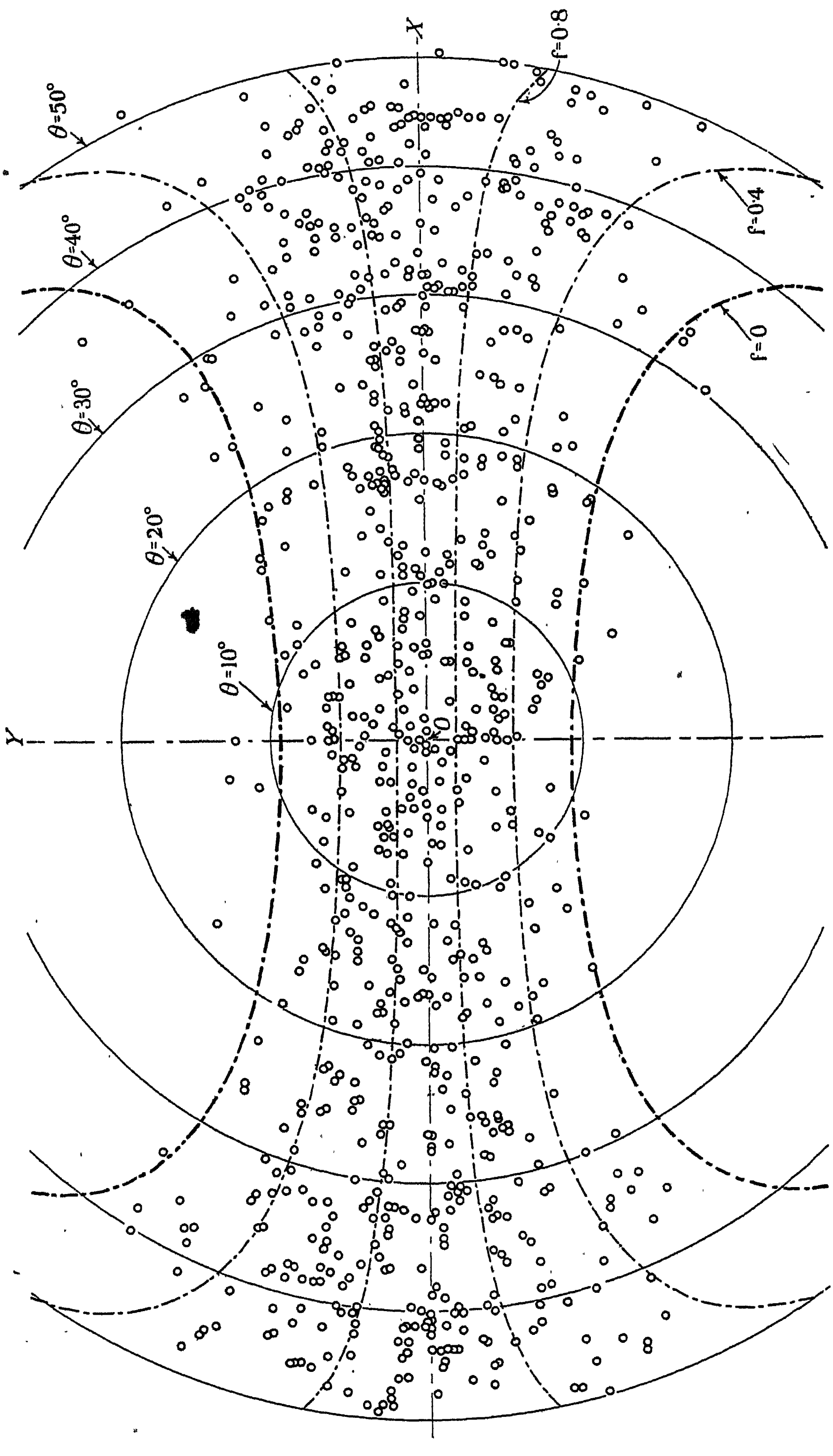


FIGURE 9. Polar diagram showing distribution in angle of projected protons produced by 8·8 MeV neutron group to illustrate effects of loss.

error. It will be seen that in fact there is a marked lack of circular symmetry, the tracks being confined to a region close to the X axis corresponding to the smaller values of β, an effect clearly due to the escape of tracks with the larger angles of dip.

It has been shewn that the fraction, f, of tracks of the class (E, L, β) which remain in the emulsion, is given by

$$f = 1 - L \sin\beta/D.$$

Suppose the neutron group to be homogeneous in energy with the value E_0. The energy E of protons projected at an angle θ is then equal to $E_0 \cos^2\theta$, and their range L can be determined. Corresponding to this value of θ, the successive values of β for which f has the values 0, 0·1, 0·2, 0·3, etc. can be determined from the equation above. Making such calculations for different values of θ, lines of constant f can be constructed, and such lines have been included in figure 9. One quadrant of the corresponding figure for the 13·0 MeV group is shown in figure 10.

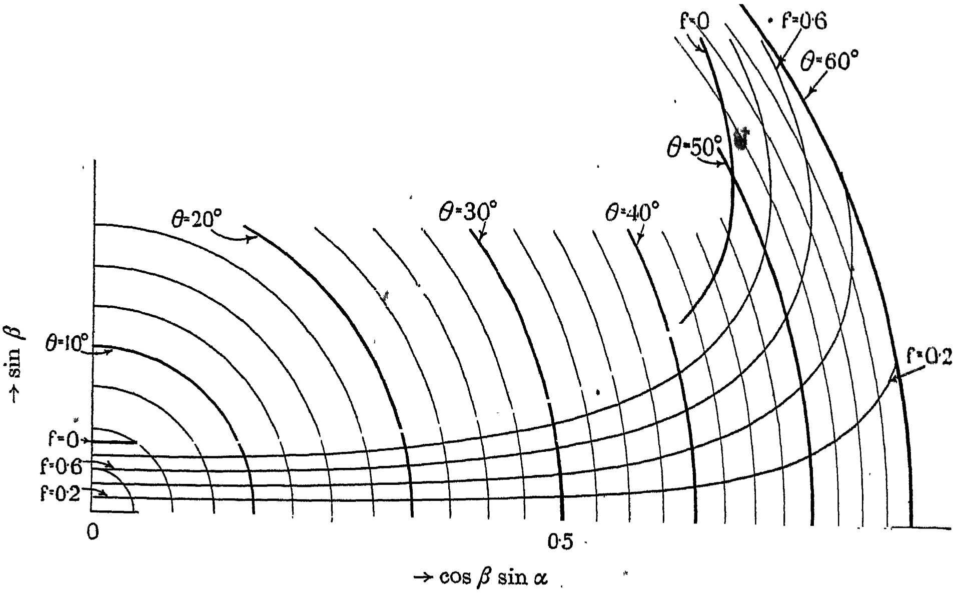


FIGURE 10. Lines of constant f and θ for 13 MeV group.

In figure 9 all but 3 % of the points lie within the region defined by the lines $f = 0$. Of the twenty tracks which lie outside this region, ten would fall within it if the observed values of β were changed by 1°. The agreement between the calculations and the observations is therefore very satisfactory in this respect, since the effects of scattering, of straggling, and of any inhomogeneity in the neutron energy in the group have been ignored.

Now consider the region of the diagram confined between two lines of constant angle of scattering, θ and $\theta + \delta\theta$. In the absence of the escape of protons the points representing the tracks would be expected to be uniformly scattered over the region, the number in any area δA being proportional to δA within the limits of statistical error. Owing to 'escape' the probability will be changed to $f\,\delta A$, where f is the mean value of f taken over the area A. The actual distributions for intervals in θ from 10 to 20°, 20 to 30°, etc. have been examined. For each interval, the areas defined between the lines $f = 0$ to $f = 0{\cdot}2$, $f = 0{\cdot}2$ to $f = 0{\cdot}4$, etc. are measured, from a large-scale drawing, by means of a planimeter. Consider a region of the diagram confined between the lines $\theta = \theta_1, \theta_2$. If, in this region, ν is the number of points in a small area δA, then $\nu/\delta A$ should be proportional to the mean value of f for the area chosen whatever its location in the region. There is consistency between the observations and the expected distribution to within the limits of the statistical error, and it can therefore be assumed that there is no serious loss in the efficiency of the observers in distinguishing tracks in the range in β involved in the measurements, i.e. from $\beta = 0°$ to $\beta = 15°$.

A similar analysis of the results for the higher angles of scattering shows that for angles of dip greater than 30° there is an increasing tendency for the observers to fail to distinguish tracks. In analysing the results to determine the variation in the intensity of scattering with angle only those tracks have been taken for which $\beta < 10°$. The proportion of measured tracks thus rejected is only 10 % because of the small values of f for the larger values of β, and by imposing this limitation it can be assured that no errors due to the failure of the observer to distinguish tracks are introduced into the results.

10. DETERMINATION OF THE SCATTERED INTENSITY

The analysis described in the previous paragraph enables one to calculate the fraction, F, of all the tracks starting in a given area of the plate, and produced by a given neutron group, which will be accepted for measurement. Consider the range of angles from θ_1 to θ_2 in the appropriate diagram of the type shown in figures 9 and 10, and let the mean number of tracks per unit area of this diagram, which would be observed in the absence of loss, be ν_θ. The total number of tracks in this range in the angle of scattering is then $\nu_\theta \pi(\sin^2\theta_2 - \sin^2\theta_1)$. Since these tracks for which $|\beta| > 10°$ are rejected, the number of tracks actually measured and accepted will equal $4\nu_\theta \int f\,dA$, where the integration is carried out in one quadrant of the diagram only, from $\theta = \theta_1$ to $\theta = \theta_2$ and from $\beta = 0$ to $\beta = 10°$. The value of F for this range in the angle of scattering, F_θ, is then given by

$$F_\theta = 4\int f\,dA/\pi(\sin^2\theta_2 - \sin^2\theta_1).$$

The values of F_θ have been calculated for the 9 and 13 MeV groups in the range from $\theta = 0°$ to $\theta = 45°$ by means of a planimeter and a large-scale drawing as described in the previous paragraph, taking intervals of 0·1 in the value of f, and of 2·5° in the value of θ. The values of F_θ thus obtained are shown for the two neutron groups of highest energy in figure 11.

Let $n_\theta \delta\theta$ be the number of measured tracks in the range from θ to $\theta + \delta\theta$, $\beta < 10°$. The total number of tracks of this type starting in the emulsion in the area examined is then given, within the limits of statistical error, by

$$N_\theta \delta\theta = n_\theta \delta\theta / F_\theta .$$

If I_θ represents the intensity of the scattered protons per unit solid angle, then $I_\theta 2\pi \sin\theta\, \delta\theta = N_\theta \delta\theta$ and $I_\theta = \kappa N_\theta \operatorname{cosec}\theta$, where κ is a constant.

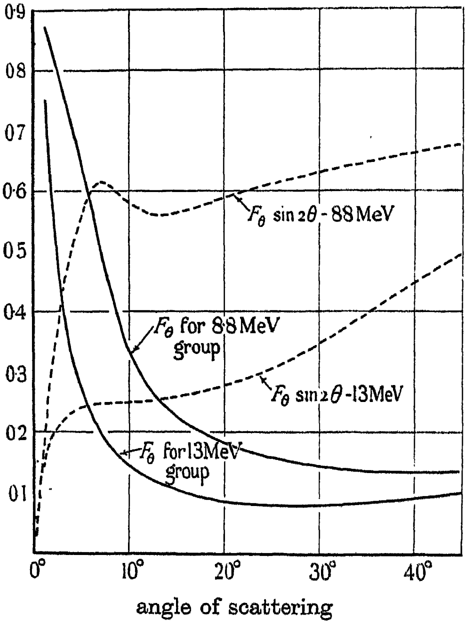


FIGURE 11. Variation of F_θ and $F_\theta \sin 2\theta$ with θ for 8·8 and 13·0 MeV groups; $F_\theta \sin 2\theta$ in arbitrary units.

Since the proton and the neutron are of nearly equal mass, a proton projected at an angle θ in the laboratory co-ordinate system has an angle of projection $\phi = 2\theta$ in the co-ordinate system of the centre of mass of the colliding particles. If I_ϕ is the intensity of scattering per unit solid angle in the centre of mass system, then

$$I_\phi \sin\phi\, \delta\phi = I_\theta \sin\theta\, \delta\theta; \quad I_\phi = I_\theta \frac{1}{4\cos\theta} = \kappa_1 N_\theta \operatorname{cosec} 2\theta .$$

It follows that if the scattering is isotropic in the centre of mass system, corresponding to a constant value for I_ϕ over the whole range of angles, then $n_\theta / F_\theta \sin 2\theta$ should also be constant, i.e. n_θ proportional to $F_\theta \sin 2\theta$. The investigated range from 0 to 45° in the angle of the projected protons corresponds to a range from 0 to 90° in the centre of mass system. The intensity of the projected protons in the centre of mass system at angle ϕ is equal to the intensity of the neutrons scattered through $\pi - \phi$ in this system. Our results therefore give the variation of the scattered intensity for the neutrons in the backward direction, i.e. from 90 to 180° in the centre of mass system.

11. Results for the 8·8 MeV group of neutrons

The results obtained for the 8·8 MeV neutron group are represented in figure 12, in which is plotted the observed numbers of tracks in successive intervals of 1° in the angle of scattering in the range from 0 to 45°. In this figure the full line represents the distribution to be expected if the scattering is isotropic. The form of this line is given by the relation $I_\theta = CF_\theta \sin 2\theta$, where the constant C is chosen to give a distribution corresponding to the number of tracks, 881, which fulfil the angular restrictions $\beta < 10°$, $\theta < 45°$, and which are identified as being produced by the 8·8 MeV group. It is a satisfactory feature of the method that for this energy isotropic scattering would produce an almost uniform distribution of tracks in the range from 0 to 45° in the angle of scattering.

It will be seen from figure 12 that the general trend of the results is in accord with the assumption of isotropic scattering. This result is displayed in table 1, where we

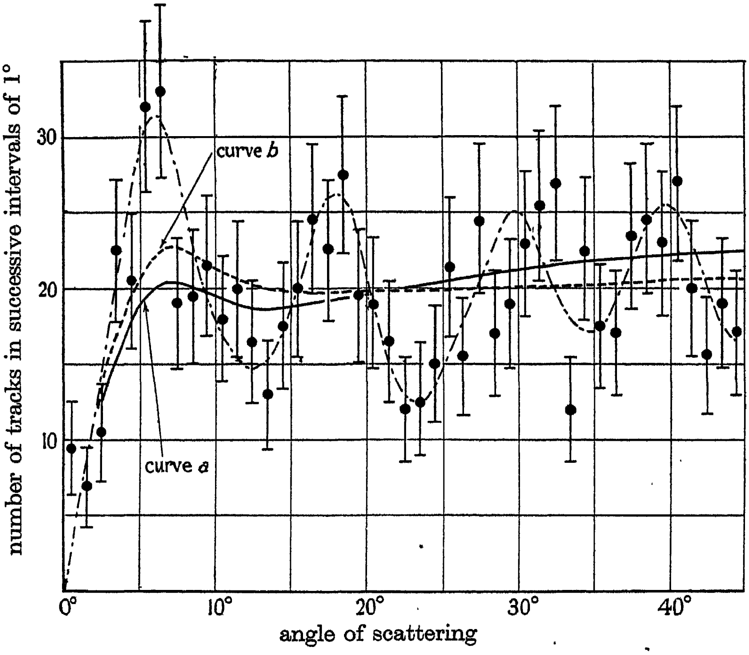


Figure 12. 8·8 MeV group. Observed distribution of projected protons in successive intervals of 1° in the angle of scattering. The full line a shows the distribution expected if scattering is isotropic. Dotted line b shows estimated best smooth curve.

Table 1

angular interval in degrees	expected number of tracks	observed number
0– 9	134	173½
9–18	171	173½
18–27	181	159
27–36	195	188
36–45	200	186½

compare the expected number of tracks in successive range intervals of 9°, with the number observed.

An alternative method of representation is shown in figure 13, where we plot, for successive intervals of 2° in the angle of scattering, the ratio of the observed number of tracks to that expected. In this diagram isotropic scattering would be represented by the horizontal line corresponding to unity in the ratio of the observed to the calculated intensity. The full line in this figure is the estimated best smooth curve through the experimental points and has a form roughly similar to that expected if there were a small *p*-component in the scattered intensity. The dotted curve *b* in figure 12 is then drawn to correspond with the full line in figure 13, and also corresponds to a total of 881 tracks.

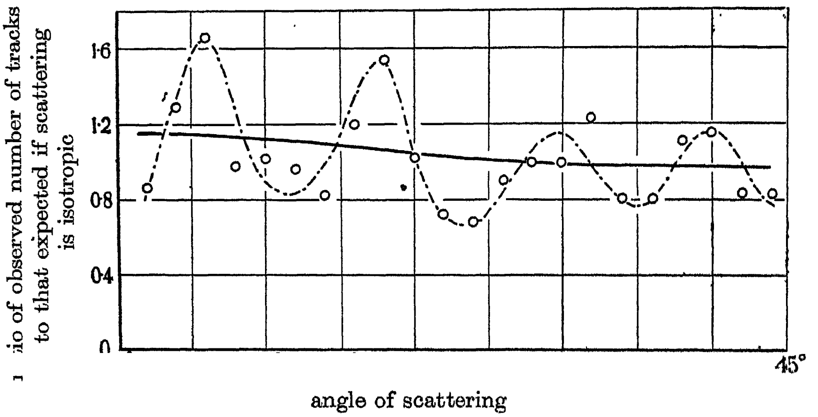


Figure 13. 8·8 MeV group. Ratio of observed number of tracks in successive intervals of 2° to that expected if scattering is isotropic. Dotted line shows best smooth curve.

An inspection of figure 12 shows that there is some indication of a periodic variation of high order in the intensity, especially in the angular range from 0 to 30°. We display these fluctuations in figure 14, where we plot for every integral value of the angle of scattering the number of tracks in angular intervals of 3°. It will be seen that there are maxima at 6, 18 and 30°, and minima at 12 and 24° in this curve. Whilst the existence of variations of this character in the scattered intensity seemed very unlikely, we have calculated the probability that the observed departures from the smooth curve shall be as great as those actually found if they are due to statistical fluctuations alone. If m is the expected value for the intensity at any angle and $m+x$ the observed value, then the quantity $\chi^2 = \Sigma x^2/m$ can be calculated for the observations. Then the level of significance of the results can be derived from the values of χ^2, calculated from a normal distribution, given by Fisher & Yates (1938). Taking the number of tracks in intervals of 1°, for the range from 1 to 25°, and taking the expected distribution to be that corresponding to the full curve in figure 13, it is found that $\chi^2 = 29{\cdot}2$. Since the dotted curve corresponds to the total number of tracks actually observed, the number of independent groups is $24-1=23$.

The table for χ^2 then indicates that the probability that the errors will be equal to or greater than those actually observed is 7 %. If a similar calculation is made grouping the tracks in successive range angular intervals of 2°, instead of 1°, an almost equal level of significance is found. The level of significance is usually fixed arbitrarily by statisticians at 5 %, so they would not regard the departure from a smooth distribution as definitely established, but physicists commonly employ a very much less stringent test of significance.

It is very difficult to account for a variation of the type suggested by these results as a consequence of deficiencies or errors in the apparatus or the method. Any broadening of the source, or any errors in the estimate of its location, would in general lead to a blurring out of the angular resolving power of the method without introducing harmonic variations of the type suggested by the results if the scattering is in fact isotropic. It is also most unlikely that the fluctuations can have been introduced through an unconscious prejudice on the part of the observers, because these features of the results were only suspected a year after the conclusion of the measurements when we found time from other duties to carry through the necessary calculations on a large part of the experimental material. Further, checks on the accuracy of the measurements of one observer by another, which are made possible by the maps of the measured areas, have never disclosed any appreciable errors in the measurements or any tendency of an observer to choose tracks with special characteristics. It therefore seems desirable to continue the measurements on plates exposed in similar conditions to those already employed in order that the statistical errors may be further reduced. If the existence of these fluctuations is then maintained, the experiment should be repeated using a neutron group of similar energy from a different source such as is provided by the bombardment of fluorine by deuterons. We suggest such an experiment because in general the methods of observation hitherto available determine only the energy and momentum

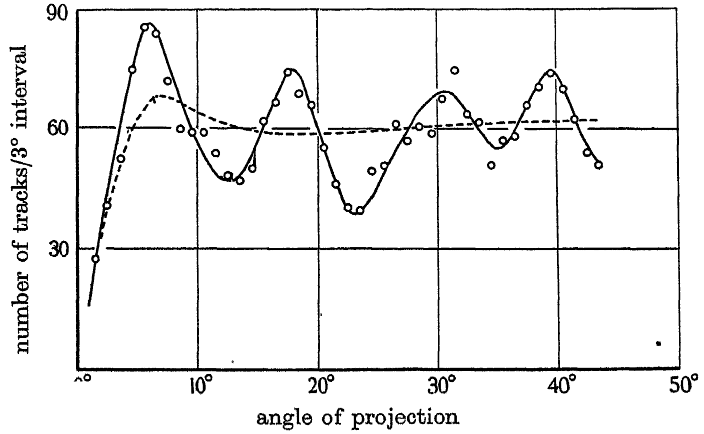


FIGURE 14. 8·8 MeV group. Number of tracks in intervals of 3° in the angle of scattering plotted for every degree.

of the fast neutrons emitted from nuclear reactions. It is unnecessary to stress the importance for nuclear theory of the variations suggested by the present experiments should they prove to be real.

12. Results for the 13 MeV group

The number of tracks produced by the 13 MeV group, which have been measured, is only 132, so that no detailed study of the distribution of the intensity of the scattered particles with angle is possible. The results are represented in figure 14 where we plot the number of tracks in successive range intervals of 5°. The calculated distribution to be expected if the scattering is in fact isotropic is shown by the full line in this figure.

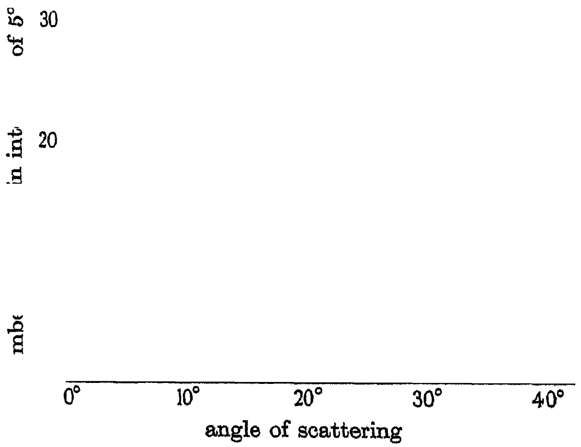


Figure 15. 13·0 MeV group. Observed distribution of scattered particles in successive intervals of 5°. Full line gives distribution expected if scattering is isotropic.

Figure 15 shows that the present results are in accord with the results of Amaldi and others to which reference has already been made. Thus our value for the ratio of the intensity of the scattered neutrons in the centre of gravity system of the colliding particles $I(\pi)/I(\pi/2)$ is $0{\cdot}67 \pm 0{\cdot}20$. This value is based on a comparison of the expected number of protons projected at angles from 0 to 10° and from 35 to 45° with those actually observed, and agrees to within the limits of the statistical error with the values ranging from 0·52 to 0·71 obtained by Amaldi for neutrons of approximately the same energy. It is clear that for this group of neutrons also it is desirable to extend the observations in order that the statistical error may be further reduced.

We are greatly indebted to Professor A. M. Tyndall for his interest in the progress of the work and for making available to us the facilities of the laboratory. Our thanks are also due to Mr H. Heitler for his collaboration in the earlier measurements, and to Miss M. Littleton for assistance in some of the calculations.

REFERENCES

Amaldi, Bocciarelli, Feratti & Trabacchi 1942 *Naturwissenschaften*, **30**, 582.
Bonner & Brubaker 1936 *Phys. Rev.* **50**, 308.
Chadwick & Goldhaber 1935 *Proc. Roy. Soc.* A, **151**, 479.
Cockcroft & Lewis 1936 *Proc. Roy. Soc.* A, **154**, 261.
Dee & Gilbert 1937 *Proc. Roy. Soc.* A, **163**, 265.
Fisher & Yates 1938 *Statistical tables for agricultural, biological and medical research.* Oliver and Boyd.
Hulthen 1943 *Phys. Rev.* **63**, 383.
Massey & Buckingham 1937 *Proc. Roy. Soc.* A, **163**, 281.
Powell & Fertel 1939 *Nature, Lond.*, **144**, 115.
Powell 1940 *Nature, Lond.*, **145**, 155.
Powell, Heitler & Champion 1940 *Nature, Lond.*, **146**, 716.
Powell 1943 *Proc. Roy. Soc.* A, **181**, 344.
Rarita & Schwinger 1941 *Phys. Rev.* **59**, 556.
Richards 1941 *Phys. Rev.* **59**, 796.
Shapiro 1941 *Rev. Mod. Phys.* **13**, 58.
Tatel 1942 *Phys. Rev.* **61**, 450.

On the equation of state for solids

BY R. FÜRTH, *Edinburgh*

(*Communicated by M. Born, F.R.S.—Received* 27 *October* 1943)

The problem of the thermodynamics of crystal lattices has been treated by rigorous methods recently in a series of papers by Born and collaborators. In particular, Bradburn succeeded in deriving the equation of state for a solid cubic crystal, consisting of identical atoms, under the assumption that the mutual potential energy of a pair of atoms satisfies a law of the form $\phi = -ar^{-m} + br^{-n}$. In the present paper a method is developed which makes it possible to determine the exponents m and n in the force law for a given element from measurements of the sublimation energy, the compressibility, the thermal expansion coefficient, and the dependence of these quantities on pressure and temperature. The method is applied to a large number of elements, and it is shown that the compression and the thermal expansion of these substances, as predicted by the theory, are in satisfactory agreement with the measured values of these quantities up to very high pressure and up to temperatures near the melting-point. The question whether melting is caused by the mechanical instability of the lattice is also investigated, and a certain rule connecting the two phenomena is found which is closely related to Lindemann's law.

1

The problem of deriving the equation of state of a solid substance from its molecular properties can be solved in two steps. First, the structure of the solid at zero temperature and pressure can be found as the equilibrium configuration of a system of particles under the action of the interatomic forces. Secondly, the change of volume for finite temperatures and pressures may be obtained by applying the methods of statistical thermodynamics. For the first task it is essential to know the correct form of the interatomic force law. Unfortunately, the calculation of the potential energy of an assembly of given particles as a function of their mutual distances is a very difficult problem, and it has been solved by quantum mechanical methods in a few cases only (see e.g. Mott & Jones 1936). If one is content, however, to assume as given the lattice structure, the lattice constant, and the lattice energy, the second task can be performed by supposing a fictitious force law, provided that it resembles the real one in some general features and can be made to fit it in a narrow region surrounding the equilibrium point; for the structure generally remains unaltered and the density changes only slightly even at high pressure and up to the melting temperature.

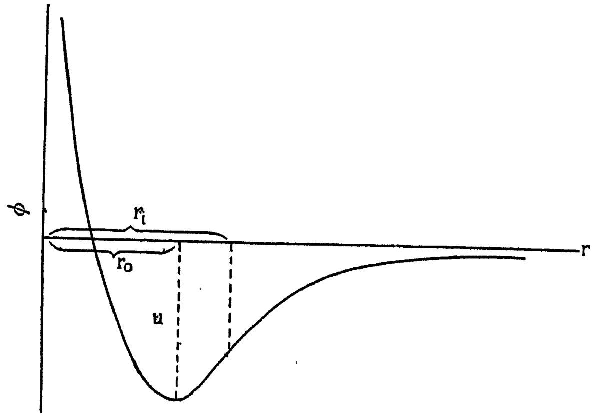


FIGURE 1. Mutual potential energy of two atoms as a function of their distance apart.

The general feature of the mutual potential energy ϕ of a pair of the constituent particles of the crystal lattice as a function of the mutual distance r between them will obviously be represented by the curve of figure 1. This curve is mainly determined by four quantities, namely, the co-ordinates of the minimum point which determine the equilibrium distance r_0 and the dissociation energy u, the curvature at the minimum point and the position of the inflexion point r_i, which determine the slope of the curve and the asymmetry in the neighbourhood of the equilibrium position. As the bulk of the lattice energy is provided by the forces between nearest neighbours, those parts of the curve which are farther away from the minimum point will not essentially affect the results as long as the general character of the curve is preserved. By the same reasoning it can be assumed that ϕ depends on the mutual distance of the particles only, even in those cases where the interatomic forces are not central as, for example, in the case of metals.

The most simple approximation of this kind is a function of the form

$$\phi = -ar^{-m} + br^{-n} \quad (m < n), \tag{1}$$

which assumes that the interatomic force is a superposition of an attracting and a repelling force, both depending on the distance r by a simple inverse power function, as is usually assumed in the kinetic theory of gases (Mie 1903). This force law also satisfies the necessary condition that the lattice held together by such forces must be mechanically stable for all kinds of small perturbations, for all values of m and n in the case of the cubic face-centred lattice, and for values of m and n smaller than 5 in the case of the body-centred cubic lattice (Born 1940; Misra 1940).

So much for the force law. The main problem in obtaining the equation of state is now to derive an expression for the free energy of the crystal lattice by the methods of statistical thermodynamics. For if this quantity is known as a function of temperature and volume the equation of state follows by known thermodynamical rules. For this purpose one has to consider the lattice, consisting of N identical particles, as a mechanical system of $3N$ degrees of freedom, capable of $3N$ normal modes of vibration about the equilibrium configuration. If all the proper frequencies are known the free energy can be calculated with the help of a well-known theorem of statistical thermodynamics.

The realization of this programme is, of course, not a simple task and attempts have been made to simplify it. One way to achieve this consists in replacing all the $3N$ proper frequencies by a single one by adopting the 'Einstein model' of a solid in which each particle is supposed to vibrate about its equilibrium position independently of the motion of the others (Grüneisen 1912), or by adopting the 'Debye model' where the solid is considered as a continuum with a continuous spectrum of proper frequencies up to an upper limit, the 'Debye frequency' (Debye 1914).

An extensive investigation was carried out by Grüneisen in order to check the results of Mie's (1903) and his own theory of the equation of state with the empirical facts, and he was able to show that certain rules connecting the sublimation energy, the compressibility, the coefficient of thermal expansion, etc., which follow from the theory were roughly satisfied by a great number of substances (see e.g. Grüneisen 1926). It must be expected, however, that a theory based on the assumption of a single characteristic frequency or of a continuous spectrum of frequencies will never yield more than a rough approximation, as these assumptions are not compatible with the crystal structure of the solid body which is an essential feature of the solid state.

The general methods for solving the problem rigorously are mainly due to Born (see e.g. Born 1923) and Stern (1916). The first attempt to derive the equation of state and expressions for the elastic constants in the limiting case of sufficiently high temperatures (i.e. high compared with the Debye temperature) is also due to Born (1939). He supposed a monatomic substance with a body-centred lattice and a force law of the type (1) with exponents $m = 6$ and $n = 12$, a combination which seemed justified both by theoretical considerations and because it had been found to fit the observations on rare gases (Lennard-Jones 1937). A comparison of the main results of this theory with the experiments on similar lines as in Grüneisen's work has been carried out elsewhere (Fürth 1941 *a*, *b*). But here again only a comparatively rough agreement could be expected, as considerable simplifications had to be introduced in the course of the theoretical calculations, and furthermore the assumption of a general force law with the above-mentioned values for the exponents for all substances could hardly be justified; also the body-centred lattice is not mechanically stable under these conditions.

The work was therefore resumed by Bradburn (Born & Bradburn 1943; Bradburn 1943), who succeeded in solving the problem rigorously for a face-centred cubic lattice; she also gave the equation of state in a numerical form for the same force law. She showed further that by making certain alterations in the averaging processes a comparatively simple formula could be derived for the equation of state of a crystal with a cubic lattice and for arbitrary exponents in the force law, which coincides practically with the rigorous result in the case of $m = 6$ and $n = 12$, so that this formula can be reasonably supposed to represent the equation of state of a large variety of substances. For a number of combinations of values of m and n sets of curves representing the isobars and isotherms according to this theory have been drawn by Gow, who has also extended the theory on the same lines to the calculation of the elastic coefficients of cubic crystals under the same conditions, and who has drawn curves showing the dependence of these quantities on temperature and pressure (Gow 1944).

With all this material available it seemed worth while to resume once again the above-mentioned investigation in order to find out whether the theoretical equation of state can be made to fit the observations and whether the constants of the force law can be determined individually for each substance in this way. The present paper is devoted to this task.

2

As mentioned in § 1 the theory is restricted to a monatomic substance where the particles are arranged in a face-centred cubic lattice. Let N be the number of atoms per mol., V the specific volume per mol., and V_0 the same quantity for zero temperature and zero pressure.

The potential energy ϕ of an isolated pair of particles of the type (1) is conveniently written in the form

$$\phi(r) = \frac{unm}{n-m}\left\{-\frac{1}{m}\left(\frac{r_0}{r}\right)^m + \frac{1}{n}\left(\frac{r_0}{r}\right)^n\right\}, \tag{2}$$

where r_0 is the equilibrium distance between the particles and u their dissociation energy.

It is further convenient to write the equation of state in a reduced form, that is, to introduce dimensionless parameters instead of the volume V, the pressure p and the temperature T by means of

$$v = V/V_0, \quad \varpi = p/p_0, \quad \tau = T/\theta, \tag{3}$$

where the constants p_0 and θ of the dimensions of pressure and temperature respectively are defined by

$$Nu = R\theta = p_0 V_0. \tag{4}$$

The equation of state can now be written in the form

$$\varpi v = F + G\tau, \tag{5}$$

where F and G are functions of volume but are independent of temperature. These functions can more easily be expressed by introducing a new parameter ξ instead of V, defined by

$$1+\xi = v^{s/3}, \tag{6}$$

where
$$s = n-m. \tag{7}$$

F and G as functions of ξ are now defined by

$$F(\xi) = a\xi(1+\xi)^{m/s}, \tag{8}$$

$$G(\xi) = c+b\frac{1+\xi}{1+K\xi}, \tag{9}$$

where a, b, c, K are constants depending only on the lattice structure and the exponents m and n, according to the following definitions:

$$a = \frac{m(s+m)}{6s}\left(\frac{S_m^0}{S_{s+m}^0}\right)^{m/s} S_m^0, \tag{10}$$

$$b = Ks/2, \tag{11}$$

$$c = \frac{m}{2}+1, \tag{12}$$

$$K = \frac{1}{1-\dfrac{m-1}{s+m-1}\, S_{s+m}^0 S_{m+2}^0 / S_m^0 S_{m+2+s}^0}. \tag{13}$$

Here the quantities S_p^0 are simple lattice sums which have been calculated and tabulated for all integer values of p between 4 and 15 for all three types of cubic Bravais lattices by Misra (1940).

With the help of the relations (5)–(13) ϖ can be calculated as a function of ξ for any pair of values of m and n, and finally from this the equation of state can be obtained in the usual way, giving the reduced volume v as a function of the reduced pressure ϖ and the reduced temperature τ. As an illustration the isobars $v(\tau)$ for different values of ϖ for the exponents $m = 6$ and $n = 12$ are shown in figure 2, taken from Bradburn's paper.

The potential energy Φ of the lattice which is equal and opposite in sign to the heat of sublimation Λ per mol. of the substance at zero temperature is given by

$$-\Phi = \Lambda = qNu, \tag{14}$$

where
$$q = \frac{1}{2}\frac{(S_m^0)^{(s+m)/s}}{(S_{s+m}^0)^{m/s}}. \tag{15}$$

From (10) and (15) follows

$$a = \frac{m(s+m)}{3s}q. \tag{16}$$

For the following calculations it is finally convenient to introduce the quantities

$$g = b+c = 1+\frac{m+Ks}{2}, \tag{17}$$

$$h = bK+c = 1+\frac{m+K^2s}{2}. \tag{18}$$

In order to carry out the proposed comparison of the theoretical equation of state with measurements it is useful first to derive formulae for the compressibility κ, the coefficient of thermal expansion α, and the dependence of these quantities on pressure and temperature. This can be done by calculating the first and second derivatives of v with respect to ϖ and τ. As under normal conditions p and T are small compared with p_0 and θ, and V is only slightly bigger than V_0, the quantities ϖ, τ, ξ are small compared with unity. The simplest procedure is therefore to expand the function $v(\varpi, \tau)$ in a power series up to the second order.

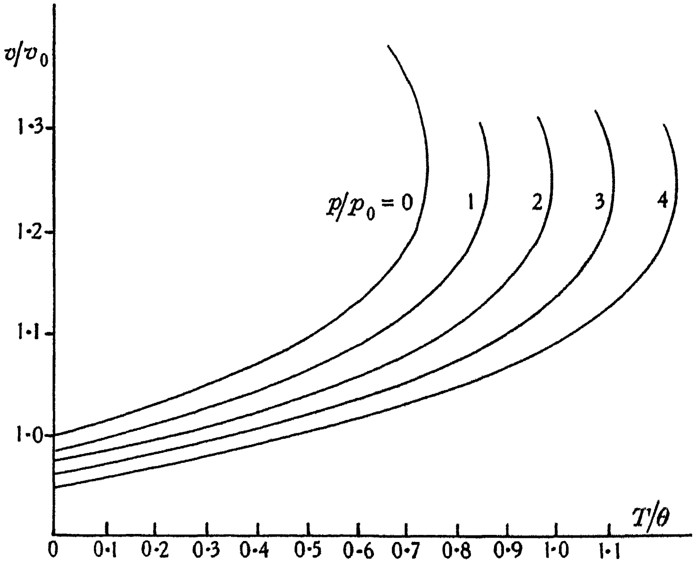


FIGURE 2. Theoretical isobars for a cubic face-centred lattice with $m = 6$, $n = 12$.

Using the relations (5), (6), (8) and (9), one obtains

$$\varpi = a\xi + \frac{a}{s}(m+3)\xi^2 + (b+c)\tau + \left[\frac{3}{s}(b+c) + b(1-K)\right]\xi\tau. \tag{19}$$

The solution of this equation with respect to ξ, to the same degree of approximation, is

$$\xi = \frac{1}{a}[\varpi - (b+c)\tau] + A\varpi^2 + B\varpi\tau + C\tau^2, \tag{20}$$

where
$$\left.\begin{aligned} A &= -\frac{m+3}{a^2 s}, \\ B &= \frac{1}{a^2}\left\{\frac{2(m+3)(b+c)}{s} - \left[\frac{3(b+c)}{s} + b(1-K)\right]\right\}, \\ C &= -\frac{b+c}{a^2}\left\{\frac{(m+3)(b+c)}{s} - \left[\frac{3(b+c)}{s} + b(1-K)\right]\right\}. \end{aligned}\right\} \tag{21}$$

From (6) and (20) one obtains further, again correct up to the second order,

$$\begin{aligned} v = 1 - \frac{3}{as}[\varpi-(b+c)\tau] - \varpi^2\left[\frac{3A}{s} - \frac{3}{2sa^2}\left(\frac{3}{s}+1\right)\right] \\ - \varpi\tau\left[\frac{3B}{s} + \frac{3}{sa^2}\left(\frac{3}{s}+1\right)(b+c)\right] - \tau^2\left[\frac{3C}{s} - \frac{3}{2sa^2}\left(\frac{3}{s}+1\right)(b+c)^2\right]. \end{aligned} \tag{22}$$

Substituting here for A, B, C from (21) one gets with the help of (17) and (18)

$$\begin{aligned} v = 1 - \frac{3}{as}[\varpi - g\tau] + \varpi^2\frac{3(9+2m+s)}{2s^2a^2} + \varpi\tau\frac{3}{sa^2}\left[h + \frac{2(3+m)}{s}g\right] \\ + \tau^2\frac{3}{sa^2}g\left[2h + \frac{3+2m-s}{s}g\right], \end{aligned} \tag{23}$$

which is the required reduced equation of state for small values of pressure and temperature. From (23) by differentiation

$$\left.\begin{aligned} \frac{\partial v}{\partial \varpi} &= -\frac{3}{as} + \frac{3(9+2m+s)}{s^2a^2}\varpi + \frac{3}{sa^2}\left[h + \frac{2(3+m)}{s}g\right]\tau, \\ \frac{\partial v}{\partial \tau} &= \frac{3g}{as} + \frac{3g}{sa^2}\left[2h + \frac{3+2m-s}{s}g\right]\tau + \frac{3}{sa^2}\left[h + \frac{2(3+m)}{s}g\right]\varpi; \end{aligned}\right\} \tag{24}$$

and hence for the limiting case $\varpi = 0$, $\tau = 0$

$$\left.\begin{aligned} &\left(\frac{\partial v}{\partial \varpi}\right)_0 = -\frac{3}{as}, \quad \left(\frac{\partial v}{\partial \tau}\right)_0 = \frac{3g}{as}, \quad \left(\frac{\partial v}{\partial \tau}\Big/\frac{\partial v}{\partial \varpi}\right)_0 = -g, \\ &\left(\frac{\partial^2 v}{\partial \varpi^2}\right)_0 = \frac{3(9+2m+s)}{s^2a^2}, \quad \left(\frac{\partial^2 v}{\partial \tau^2}\right)_0 = \frac{3g}{sa^2}\left[2h + \frac{3+2m-s}{s}g\right], \\ &\left(\frac{\partial^2 v}{\partial \varpi\,\partial \tau}\right)_0 = \frac{3}{sa^2}\left[h + \frac{2(3+m)}{s}g\right]. \end{aligned}\right\} \tag{25}$$

The compressibility κ, the thermal expansion coefficient α and their derivatives with respect to pressure and temperature can, in virtue of (3) and (4), be expressed in terms of the above quantities as follows:

$$\left.\begin{aligned} &\kappa_0 = -\frac{1}{V_0}\left(\frac{\partial V}{\partial p}\right)_0 = -\frac{1}{p_0}\left(\frac{\partial v}{\partial \varpi}\right)_0, \quad \alpha_0 = \frac{1}{V_0}\left(\frac{\partial V}{\partial T}\right)_0 = \frac{1}{\theta}\left(\frac{\partial v}{\partial \tau}\right)_0, \\ &\frac{1}{\kappa_0^2}\left(\frac{\partial \kappa}{\partial p}\right)_0 = -\left[\frac{\partial^2 v}{\partial \varpi^2}\Big/\left(\frac{\partial v}{\partial \varpi}\right)^2\right]_0, \quad \frac{1}{\alpha_0^2}\left(\frac{\partial \alpha}{\partial T}\right)_0 = \left[\frac{\partial^2 v}{\partial \tau^2}\Big/\left(\frac{\partial v}{\partial \tau}\right)^2\right]_0, \\ &\frac{1}{\alpha_0\kappa_0}\left(\frac{\partial \kappa}{\partial T}\right)_0 = -\frac{1}{\alpha_0\kappa_0}\left(\frac{\partial \alpha}{\partial p}\right)_0 = -\left[\frac{\partial^2 v}{\partial \varpi\,\partial \tau}\Big/\left(\frac{\partial v}{\partial \varpi}\frac{\partial v}{\partial \tau}\right)\right]_0. \end{aligned}\right\} \tag{26}$$

Substituting the expressions (25) in (26) one finally obtains the following relations:

$$p_0\kappa_0 = 3/sa, \tag{27}$$

$$\theta\alpha_0 = 3g/sa, \tag{28}$$

$$V_0\alpha_0/R\kappa_0 = g, \tag{29}$$

$$\frac{1}{\kappa_0^2}\left(\frac{\partial\kappa}{\partial p}\right)_0 = -\frac{9+2m+s}{3} = -\frac{9+m+n}{3}, \tag{30}$$

$$\frac{1}{\alpha_0^2}\left(\frac{\partial\alpha}{\partial T}\right)_0 = \frac{1}{3}\left[\frac{2sh}{g}+(3+2m-s)\right], \tag{31}$$

$$\frac{1}{\alpha_0\kappa_0}\left(\frac{\partial\kappa}{\partial T}\right)_0 = -\frac{1}{\alpha_0\kappa_0}\left(\frac{\partial\alpha}{\partial p}\right)_0 = \frac{1}{3}\left[\frac{sh}{g}+2(3+m)\right]. \tag{32}$$

Equation (30) has been derived by Grüneisen (1912) from his theory, and later independently by Bridgman (1923). More recently its correctness has been proved rigorously for the cubic face-centred lattice at zero temperature (Fürth 1942).

It has been further shown by Grüneisen (1912) on the basis of his theory that the quantity

$$\gamma = \frac{V\alpha}{c_v\kappa}, \tag{33}$$

where c_v is the specific heat at constant volume, should be independent of temperature for each substance; it is therefore usually called the 'Grüneisen constant'. Its value should depend on the values of the exponents in the atomic force law. It has been especially suggested by Grüneisen (1926) that γ might depend only on the exponent n of the repulsive force according to the simple formula

$$\gamma = \frac{n+2}{6}. \tag{34}$$

The calculation of γ from (33) by means of the measured values of the quantities on the right-hand side of this formula for a large variety of elements (Grüneisen 1926) gave the result that γ is not only constant for each substance separately but has also roughly the same value ($\sim 2{\cdot}4$) for all the elements with a few exceptions. This would, according to (34), lead to the universal value $n = 12$ for the exponent of the repulsive force which is therefore used in many theoretical investigations (see e.g. Lennard-Jones 1937; Born 1939).

As the present theory is restricted to sufficiently high temperatures where the classical formula

$$c_v = 3R \tag{35}$$

is valid, (33) becomes identical with (29) by putting

$$\gamma = g/3. \tag{36}$$

As, however, g depends on m as well as on n, the use of formula (34) for calculating n becomes unjustified and also the conclusion that the repulsive force has nearly the same character for most substances.

A relation similar to equation (32) has also been derived by Grüneisen (1912), namely,

$$\frac{1}{\alpha\kappa}\frac{\partial\kappa}{\partial T} = \tfrac{1}{3}(m+n+6), \tag{37}$$

which coincides with the correct expression (32) only for $m = 3$.

The right-hand sides of the equations (27)–(32) depend only on the exponents m and n and can therefore be calculated for any pair of values of m and n with the help of the formulae (10), (13), (17), (18) and Misra's tables of the lattice sums S_p^0. The results of this calculation are contained in table 1. The lattice sums are finite only for values of p smaller than 3; nevertheless, it can be seen directly that the expression on the right-hand side of (30) remains finite for all finite values of m and n and that it is a linear function of these quantities. The numerical calculation shows further that the same applies to the quantities on the right-hand sides of (29), (31) and (32), and that they are practically linear functions of s for any fixed value of m or n inside the range accessible to numerical calculation. It is therefore easy to extrapolate the values of these quantities for values of m and n smaller than 3. The expressions (27) and (28), however, vanish for $m = 3$ and are not defined for $m < 3$. They also contain the unknown quantities p_0 and θ.

In order to obtain relations for comparing the theoretical and the measured values of the thermal expansion coefficient and the compressibility one can eliminate the quantities p_0 and θ from (27) and (28) with the help of (4) and (14), introducing instead the sublimation energy Λ. One has

$$\Lambda = qR\theta = qp_0V_0, \tag{38}$$

which combined with (27) gives
$$\kappa_0\Lambda = \frac{3qV_0}{sa}, \tag{39}$$

which again by means of (16) can be transformed into

$$\kappa_0\Lambda = \frac{9V_0}{m(s+m)} = \frac{9V_0}{mn}. \tag{40}$$

This relation has been derived by Grüneisen (1912) and by Born & Landé (1918), and it is generally valid for all values of m and n.

Similarly one obtains instead of (28)

$$\alpha_0\Lambda = \frac{3gqR}{sa}, \tag{41}$$

which by means of (16) is transformed into

$$\alpha_0\Lambda = \frac{9gR}{m(s+m)} = \frac{9gR}{mn}. \tag{42}$$

This formula, of course, provides no new information, as it is a direct consequence of (29) and (40).

TABLE 1

m	s	$p_0\kappa_0$	$\theta\alpha_0$	$g=\dfrac{V_0\alpha_0}{R\kappa_0}$	$\dfrac{1}{\alpha_0^2}\left(\dfrac{\partial\alpha}{\partial T}\right)_0$	$\dfrac{1}{\alpha_0\kappa_0}\left(\dfrac{\partial\kappa}{\partial T}\right)_0$	Λ/T_c	V_c/V_0
4	1	0·00705	0·0288	4·0905	4·2199	5·1049	11·49	1·405
4	2	0·00963	0·0446	4·6297	4·6289	5·4811	14·49	1·299
4	3	0·01080	0·0557	5·1579	5·0336	5·8501	17·28	1·238
4	4	0·01122	0·0637	5·6812	5·4286	6·2143	19·94	1·198
4	5	0·01120	0·0695	6·2010	5·8157	6·5412	22·54	1·170
4	6	0·01098	0·0738	6·7173	6·1959	6·9313	25·06	1·148
4	7	0·01063	0·0769	7·2304	6·5699	7·2850	27·53	1·131
4	8	0·01024	0·0793	7·7420	6·9392	7·6363	29·97	1·119
4	9	0·00983	0·0811	8·2520	7·3047	7·9857	32·38	1·108
4	10	0·00943	0·0826	8·7615	7·667	8·334	34·77	1·098
5	1	0·01586	0·0830	5·2307	5·2095	5·9381	13·52	1·280
5	2	0·01667	0·0958	5·7799	5·6731	6·3365	15·97	1·221
5	3	0·01658	0·1048	6·3189	6·1178	6·7256	18·27	1·184
5	4	0·01608	0·1102	6·8506	6·5475	7·1071	20·46	1·158
5	5	0·01539	0·1135	7·3757	6·9639	7·4819	22·57	1·140
5	6	0·01465	0·1157	7·8980	7·3713	7·8523	24·63	1·125
5	7	0·01391	0·1171	8·4161	7·7695	8·2181	26·65	1·113
5	8	0·01320	0·1179	8·9300	8·1610	8·5805	28·62	1·103
5	9	0·01255	0·1186	9·4465	8·5473	8·940	30·60	1·095
6	1	0·01849	0·1178	6·3730	6·2631	6·7982	15·01	1·208
6	2	0·01803	0·1248	6·9239	6·7499	7·2083	17·05	1·174
6	3	0·01723	0·1286	7·4641	7·2153	7·6077	18·96	1·151
6	4	0·01631	0·1304	7·9952	7·6627	7·9980	20·77	1·133
6	5	0·01540	0·1312	8·5197	8·0952	8·3809	22·53	1·121
6	6	0·01452	0·1313	9·0420	8·5178	8·7589	24·24	1·108
6	7	0·01370	0·1310	9·5600	8·9305	9·1319	25·92	1·099
6	8	0·01294	0·1304	10·081	9·340	9·503	27·59	1·092
7	1	0·01785	0·1339	7·5012	7·3342	7·6671	16·07	1·164
7	2	0·01691	0·1361	8·0473	7·8304	8·0819	17·77	1·144
7	3	0·01591	0·1366	8·5849	8·3065	8·4866	19·38	1·129
7	4	0·01493	0·1360	9·1114	8·7621	8·8810	20·91	1·116
7	5	0·01402	0·1351	9·6345	9·2052	9·2693	22·39	1·106
7	6	0·01318	0·1338	10·154	9·6367	9·6517	23·85	1·100
7	7	0·01241	0·1325	10·678	10·065	10·033	25·31	1·090
8	1	0·01606	0·1383	8·6127	8·4075	8·5371	16·83	1·140
8	2	0·01502	0·1374	9·1511	8·9057	8·9529	18·26	1·125
8	3	0·01405	0·1360	9·6801	9·3833	9·3583	19·62	1·112
8	4	0·01315	0·1342	10·204	9·8454	9·7560	20·93	1·103
8	5	0·01234	0·1323	10·724	10·295	10·147	22·21	1·093
8	6	0·01159	0·1304	11·251	10·742	10·537	23·51	1·090
9	1	0·01404	0·1362	9·7016	9·4708	9·4021	17·34	1·110
9	2	0·01308	0·1338	10·233	9·9683	9·8175	18·56	1·110
9	3	0·01222	0·1315	10·757	10·448	10·224	19·74	1·100
9	4	0·01144	0·1290	11·277	10·913	10·623	20·89	1·092
9	5	0·01073	0·1277	11·807	11·378	11·022	22·06	1·075
10	1	0·01215	0·1309	10·771	10·522	10·261	17·67	1·100
10	2	0·01131	0·1276	11·285	11·009	10·671	18·70	1·100
10	3	0·01058	0·1249	11·807	11·491	11·079	19·74	1·090
10	4	0·00991	0·1224	12·351	11·983	11·492	20·86	1·082
11	1	0·01051	0·1243	11·824	11·563	11·115	17·90	1·095
11	2	0·00979	0·1208	12·341	12·055	11·527	18·83	1·085
11	3	0·00917	0·1183	12·896	12·571	11·952	19·89	1·082
12	1	0·00910	0·1171	12·868	12·598	11·965	18·05	1·082
12	2	0·00842	0·1132	13·443	13·145	12·406	19·10	1·080

3

As already pointed out at the end of § 1 it is the aim of this investigation to ascertain whether a force law of the type (1) is compatible with the properties of solid elements. The procedure adopted was first to derive a value for the product mn of the exponents in the force law for a particular substance from the observed values of κ_0, V_0, and S by means of formula (40). If then the sum $m+n$ can be calculated in a similar way, both exponents can be determined separately.

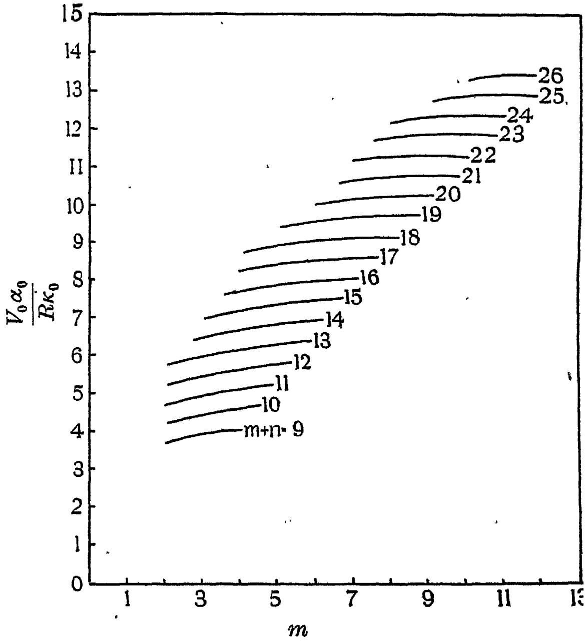


FIGURE 3. Diagram for the determination of $m+n$ from specific volume, compressibility and thermal expansion.

The most direct way of doing so is to use formula (30) by applying it to measurements of the dependence of compressibility on pressure. Further material can be obtained by using the relations (29), (31) and (32). For this purpose sets of curves were drawn showing the right-hand sides of these expressions as functions of m for integer values of $m+n$ by using the figures of table 1 and the extrapolations of these figures mentioned in § 2. The corresponding graphs are shown in figures 3–5. With the help of these graphs one can now derive values for $m+n$ from measurements of α and its dependence on temperature, and further from the dependence of κ on temperature in such a way that they are consistent with the already determined values of mn.

Table 2 contains the experimental data required for carrying out this task for all the elements for which reliable measurements were available. The values of V, the

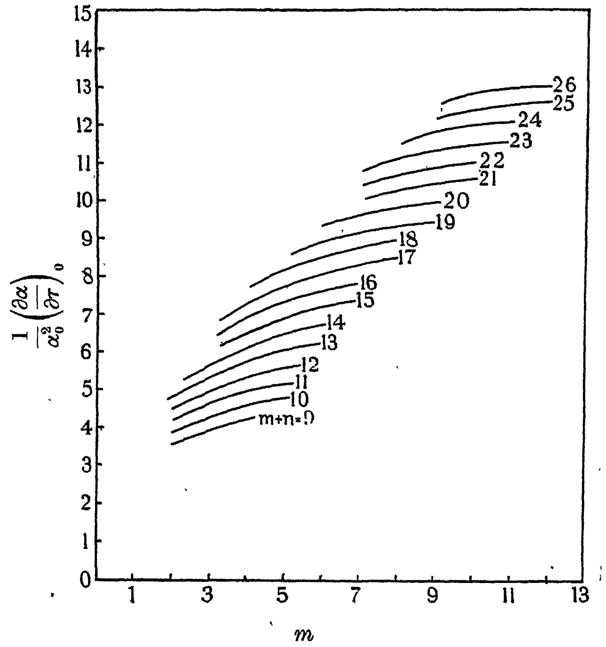


Figure 4. Diagram for the determination of $m+n$ from the temperature dependence of the thermal expansion coefficient.

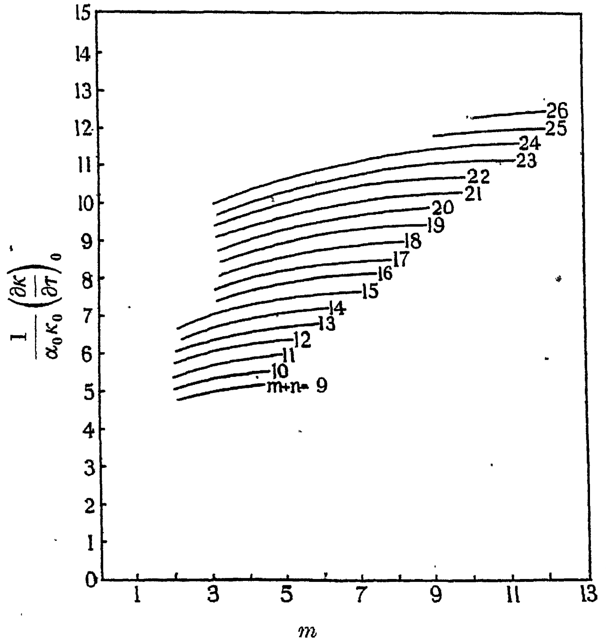


Figure 5. Diagram for the determination of $m+n$ from the dependence of compressibility on temperature.

molar volume, and those of Λ, the sublimation heat per mol., are taken from the Landolt-Börnstein tables (Roth & Scheel 1923–36). The values of κ, the compressibility at low pressure, are taken from the extensive measurements of Bridgman (1931, 1938, 1940, 1941). The limiting values of $\partial\kappa/\partial p$ were also derived from these measurements by graphical extrapolation for the more compressible substances and by using the two constant empirical formula of Bridgman for the less compressible ones; these constants had to be corrected according to a remark by Slater (1940). The temperature coefficient of compressibility $\frac{1}{\kappa}\frac{\partial\kappa}{\partial T}$ is also found from Bridgman's experiments. Assuming that this coefficient is independent of pressure one sees easily that the temperature dependence of $\frac{1}{\kappa^2}\frac{\partial\kappa}{\partial p}$ is the same as that of $1/\kappa$; this shows the procedure for obtaining the limiting values of $\frac{1}{\kappa^2}\frac{\partial\kappa}{\partial p}$ for zero temperature and pressure from the data mentioned above; they are also tabulated in column 11 of table 2.

As the present theory is based on classical statistics and consequently only holds for temperatures above the Debye temperature, where quantum effects can be neglected, the limiting values α_0 of the thermal expansion coefficient are not the actual values of this quantity at zero temperature (which according to Grüneisen's well-known law should be infinitely small), but such values as can be extrapolated for zero temperature from measurements of the dependence of α on temperature in the region above the Debye temperature. For all substances for which measurements of this kind were available the limiting values of β_0, the coefficient of linear thermal expansion, have been extrapolated and tabulated in column 7 of table 2. In order to get reliable values of the quantity $\frac{1}{\alpha_0}\left(\frac{\partial\alpha}{\partial T}\right)_0$ in the same way the calculations must be based on extremely accurate observations. Only the experiments of a few of the many experimenters in this field (Austin 1932; Austin & Pierce 1933; Eucken & Dannöhl 1934; Nix & MacNair 1941, 1942) were found to satisfy this condition. Values of $\frac{1}{\alpha_0}\left(\frac{\partial\alpha}{\partial T}\right)_0$ were calculated by numerical and graphical extrapolation from the data given by these authors; for substances where measurements had been made by different observers the averages of the respective results were taken. The final values are given in column 8 of table 2; they, together with the corresponding values of $\beta_0 = \frac{1}{3}\alpha_0$ in column 7, were eventually used for calculating the values of $\frac{1}{\alpha_0^2}\left(\frac{\partial\alpha}{\partial T}\right)_0$ in column 12.

The values of $\frac{1}{\alpha_0\kappa_0}\left(\frac{\partial\kappa}{\partial T}\right)_0$, tabulated in column 13 of table 2, are partly calculated from the corresponding values in the preceding column and partly from the special measurements of Grüneisen (1910). These figures show that the thermal coefficient of κ is in the average about five times bigger than α, the thermal coefficient of V. As Λ decreases with increasing temperature and thus has a negative temperature

TABLE 2

element	V (cm.^{3}/mol.)	Λ (kcal./mol.)	κ (cm.^{2}/kg.) $\times 10^6$	$-\left(\frac{\partial\kappa}{\partial p}\right)_{p=0}$ $\times 10^{12}$	$\frac{1}{\kappa}\frac{\partial\kappa}{\partial T}$ $\times 10^4$	$\beta_0\times 10^6$	$\left(\frac{1}{\alpha}\frac{\partial\alpha}{\partial T}\right)_0$ $\times 10^4$	$\frac{9V_0}{\kappa_0\Lambda}$	$\frac{V_0\alpha_0}{R\kappa_0}$	$-\frac{1}{\kappa_0^2}\left(\frac{\partial\kappa}{\partial p}\right)_0$	$\frac{1}{\alpha_0^2}\left(\frac{\partial\alpha}{\partial T}\right)_0$	$\frac{1}{\alpha_0\kappa_0}\left(\frac{\partial\kappa}{\partial T}\right)_0$
Ne	13·9	0·6	—	—	—	—	—	—	—	—	—	—
A	23·6	2·05	—	—	—	—	—	—	—	—	—	—
Kr	—	(2·8)	—	—	—	—	—	—	—	—	—	—
X	—	(4·1)	—	—	—	—	—	—	—	—	—	—
Li	13·0	36·0	8·7	(400)	7·1	56	—	8·8	2·9	6·3	—	4·2
Na	23·7	26·2	15·6	950	9·6	71	—	12·2	3·8	5·0	—	4·5
K	45·5	21·8	35	6,000	14·0	83	—	12·6	3·8	6·0	—	5·6
Rb	56·0	20·6	51	12,000	—	90	—	11·2	3·5	(6·0)	—	—
Cs	71·0	18·7	67	24,000	—	97	—	12·0	3·6	(8·0)	—	—
Cu	7·17	81·7	0·72	2·8	4·6	14·7	2·51	26	5·6	6·2	5·7	8·9
Ag	10·3	69·4	0·99	6·2	12·8	17·4	2·40	32	6·9	8·9	4·6	7·3
Au	10·2	90·7	0·57	2·6	—	13·3	2·50	42	8·8	(12)	6·5	—
Be	4·89	55·0	0·80	4·8	2·8	9·2	—	24	3·3	8·1	—	6·2
Mg	13·9	36·6	2·95	54	—	24·8	—	27·2	4·2	(7)	—	—
Ca	25·8	42·9	5·70	180	5·9	30	—	22	4·8	6·5	—	6·6
Sr	33·7	40·0	8·19	360	2·2	—	—	22	—	4·7	—	—
Ba	48·3	40·9	10·2	380	9·6	—	—	19·4	—	4·6	—	—
Zn	8·83	31·4	1·6	7·2	—	28·3	—	37	5·5	(3·5)	—	—
Cd	13·1	27·0	2·41	(25)	5·6	31·7	—	42	4·9	(5)	—	(7)
Hg	14·8	15·5	3·9	—	—	42	—	51	5·6	(5)	—	—
B	6·25	(80)	0·55	1·8	—	8·3	—	(30)	3·3	(7)	—	—
Al	10·0	67·6	1·34	9·0	6·3	21·6	(5·5)	24	5·7	6·0	(8·5)	9·1
Tl	17·2	43·5	2·4	(30)	(12·1)	29·4	—	34	7·4	(7)	—	—
C	3·42	(199)	0·18	—	—	1·3	—	(27)	3·3	—	—	—
Si	12·6	94·3	0·3	—	—	3·6	—	92	5·2	—	—	—
Sn	16·3	70	1·70	15	4·7	16·8	—	29	5·7	5·9	—	9·3
Pb	18·2	47	2·31	(30)	5·6	24·3	(7·0)	36	8·1	(6·7)	(9·5)	6·4
As	13·1	31	5·6	180	—	5·3	—	16	(0·4)	(6·2)	—	—
Sb	18·2	54·4	2·6	(35)	5·4	9·0	—	27	(2·2)	(5·9)	—	(20)
Bi	21·4	—	2·9	52	4·8	13·4	—	—	(3·5)	7·1	—	(12)
Cr	7·32	89·5	0·52	1·7	5·2	6·7	—	33	3·3	(7)	—	(39)
Mo	9·40	156	0·35	—	0·64	4·5	(2·2)	37	4·8	—	(16)	4·3
W	9·64	203·9	0·4	2·4	1·5	3·5	(11·4)	34	5·2	5	(110)	11·3
Se	16·5	28	9·8	450	8·7	—	—	13	—	5·9	—	—
Mn	7·55	69	0·80	8	4·7	21	—	29	7·1	(14·5)	—	7·6
Fe	7·16	96·5	0·59	1·6	2·3	12·3	—	27	5·3	5·0	—	9·5
Co	6·70	—	0·54	1·6	3·3	12·6	—	—	5·5	6·1	—	8·7
Ni	6·68	98	0·53	1·6	1·3	11·6	4·3	27	6·1	6·0	12·5	6·0
Pt	9·13	124·7	0·36	1·0	2·5	8·5	1·2	43	7·9	8·5	4·6	8·6

coefficient, it can be reasonably assumed (within the accuracy of this investigation) that the temperature coefficient of $V/\kappa\Lambda$ is very small. Consequently this quantity has been calculated from the measurements of V, Λ, and κ at room temperature contained in the columns 2, 3, and 4; the results are shown in column 9. Since, finally, the temperature coefficients of α and κ are of the same order of magnitude, and that of V, as already mentioned, about five times smaller, the quantity $V\alpha/\kappa$ can also be considered as being practically independent of temperature; the figures given in column 10 of table 2 are calculated on this assumption.

Table 3 contains the values of mn and $m+n$, calculated from the data of table 2 in the way indicated above, and finally, in the last two columns, the probable values of the exponents m and n separately. The less reliable figures are bracketed. The values for the rare gases are quoted according to Lennard-Jones (1937), derived from measurements on the thermal properties of these substances in the gaseous state near the critical point.

One sees that substances belonging to the same group of the periodic system mostly have the same or very similar values of m and n. But the variation of the values throughout is not very wide except in a few cases. This accounts for the fact that quite reasonable results had been obtained in former papers by the present author (Fürth 1941 *a*, *b*) where he had tried to account for the mechanical and thermodynamical properties of solids on the basis of Born's theory in its first form (Born 1939) where fixed values of the exponents ($m = 6$ and $n = 12$) had been used.

4

The fact that consistent values of the exponents m and n in the atomic force law can be derived from measurements at low pressure and temperature for a large number of substances can be regarded as evidence for the correctness of the theory. Further evidence of a still more convincing character can be obtained by comparing the measured dependence of volume on temperature and pressure over a wide range with the dependence predicted by the equation of state (5) in connexion with equations (6)–(12), when the above-mentioned values of m and n are substituted. Ideal experimental material for this programme is provided by recent measurements of Bridgman on the compression of elements under very high pressure up to 100,000 atm. (Bridgman 1940, 1941).

In order to carry out this task it is convenient to write equation (5) by using (3), (16) and (38) in the form

$$p = \frac{1}{V_0 v}[\Lambda H(\xi) + RT\, G(\xi)] \tag{43}$$

with
$$H(\xi) = \frac{m(s+m)}{3s}\,\xi(1+\xi)^{m/s}. \tag{44}$$

Writing further
$$\delta V = V - V_0 \tag{45}$$

for the difference between the actual volume V and the volume V_0 at zero pressure and zero temperature, one has further

$$v = 1+\delta V/V_0 \tag{46}$$

and

$$\xi = (1+\delta V/V_0)^{s/3}-1. \tag{47}$$

By using these equations together with the equations (9), (11), (12) and (13), one can now calculate p as a function of δV for any temperature T if m, n, V_0, and Λ are given.

TABLE 3

element	$m \times n$ from $\dfrac{9V_0}{\kappa_0 \Lambda}$	$m+n$ from $\dfrac{V_0\alpha_0}{R\kappa_0}$	$m+n$ from $-\dfrac{1}{\kappa_0^2}\left(\dfrac{\partial\kappa}{\partial p}\right)_0$	$m+n$ from $\dfrac{1}{\alpha_0^2}\left(\dfrac{\partial\alpha}{\partial T}\right)_0$	$m+n$ from $\dfrac{1}{\alpha_0\kappa_0}\left(\dfrac{\partial\kappa}{\partial T}\right)_0$	m	n
Ne	—	—	—	—	—	6	12
A	—	—	—	—	—	6	12
Kr	—	—	—	—	—	6	12
Xe	—	—	—	—	—	6	12
Li	9	7–8	10	—	8	1·5	6
Na	12	9	6	—	9	2	6
K	12·5	9	9	—	12	2	6
Rb	11	8–9	(9)	—	—	2	6
Cs	12	8–9	(15)	—	—	2	6
Cu	26	12–13	9	13	18–22	4	7
Ag	32	14–15	17	11	15–18	4·5	7
Au	42	18–19	(27)	13	—	5·5	8
Be	24	8	15	—	12–14	4	6
Mg	27	9–10	(12)	—	—	4	6
Ca	22	11	10	—	13–15	4	6
Sr	22	—	5	—	—	(4)	(6)
Ba	19	—	5	—	—	(4)	(6)
Zn	37	12–13	(2)	—	—	5	7
Cd	42	11	(6)	—	(14–17)	6	7
Hg	51	12–13	(6)	—	—	6	8
B	(30)	8	(12)	—	—	(4)	(7)
Al	24	12–13	9	(19)	19–24	4	7
Tl	34	15–16	(12)	—	—	4	9
C	27	8–9	—	—	—	(4)	(6)
Si	92	11–12	—	—	—	(8)	(12)
Sn	29	12–13	9	—	19–24	3	10
Pb	36	16–17	(11)	(21)	13–15	3	12
As	16	?	(10)	—	—	(2)	(8)
Sb	27	(5–6)	(9)	—	?	(4)	(7)
Bi	—	(8–9)	12	—	?	(4)	(7)
Cr	33	8	(12)	—	?	5	7
Mo	37	11	—	?	8	5	7
W	34	11–12	6	?	?	5	7
Se	13	—	9	—	—	3	5
Mn	29	14–15	(34)	—	15–18	3	10
Fe	27	11–12	6	—	19–22	4	7
Co	—	12–13	9	—	18–22	4	7
Ni	27	13–14	9	18–22	11–13	4	7
Pt	43	16–17	16	11	18–22	5·5	8

The quantity measured and tabulated in Bridgman's publications is the actual compression ΔV, that is, the difference between the volume $\bar{V}$ at zero pressure and the volume V at pressure p, both at the same temperature T, divided by V, which, as the difference between V and V_0 is small compared with V_0, can be written in the form

$$\Delta V = \frac{\bar{V}-V}{V} \approx \frac{\bar{V}-V}{V_0}$$

or

$$\Delta V = \frac{\delta\bar{V}}{V_0} - \frac{\delta V}{V_0}, \tag{48}$$

where

$$\delta\bar{V} = \bar{V} - V_0. \tag{49}$$

As according to (45) and (49) $\delta V = \delta\bar{V}$ for $p = 0$, $\delta\bar{V}$ can be easily found from the above calculations as the value of δV for which p becomes zero. This makes it possible, by means of (49), to calculate p as a function of ΔV, and so to draw a curve representing the theoretical dependence of ΔV on p which can be directly compared with the experimental results. Finally, the theoretical value of $\delta\bar{V}$ can be compared with the experimental value of this quantity derived from observations on the thermal expansion at zero pressure.

The calculations were carried out for all substances for which values of the exponents m and n had been determined and for which, at the same time, measurements of the compression up to sufficiently high pressures had been performed without showing any indication of a 'pressure transformation' of the lattice. The results are shown in figures 6–10 and in table 4. It can be stated that the agreement between theory and experiment is in general very satisfactory as regards the compressions up to the highest pressures at different temperatures, and the thermal expansion at zero pressure.

5

Figure 2 shows that each isobar extends up to a certain critical temperature T_c where it turns over. This indicates that by reaching this temperature the lattice becomes unstable, and hence the upper branch of the isobar has no real significance. One is tempted to identify T_c with the melting-point T_m and so to explain the process of fusion as the breaking up of the lattice by the thermal movement. This idea was first expressed by Herzfeld & Goeppert-Mayer (1934), and later Born derived the general conditions for the stability of a crystal lattice and showed that melting criteria could be expressed in the form of certain conditions for the elastic constants, one of which is the vanishing of the rigidity modulus: $c_{44} = 0$ (Born 1939). In the cases investigated numerically by Gow (1944), the sharpest of these instability criteria always practically coincides with the criterion $T_m = T_c$.

There are, however, two strong objections against this simple explanation of melting. First, if it were correct, the coefficient of thermal expansion α, which is given by the slope of the isobars, would increase very rapidly on approaching the

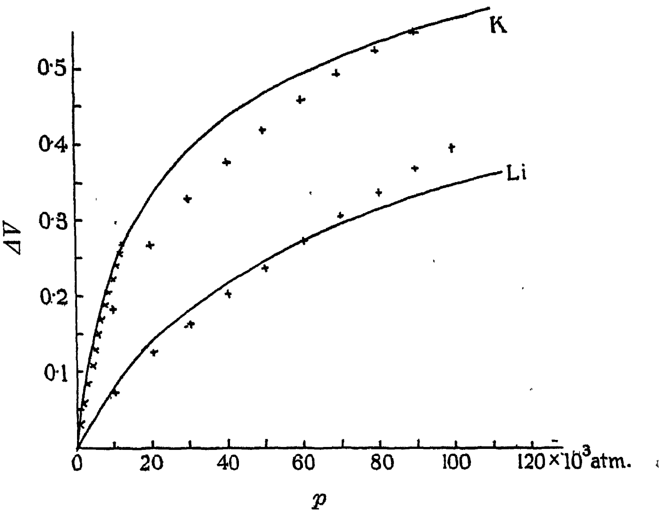


FIGURE 6. Compression curves of K and Li (× measured by Bridgman 1923, + measured by Bridgman 1941). $T = 300°$ K.

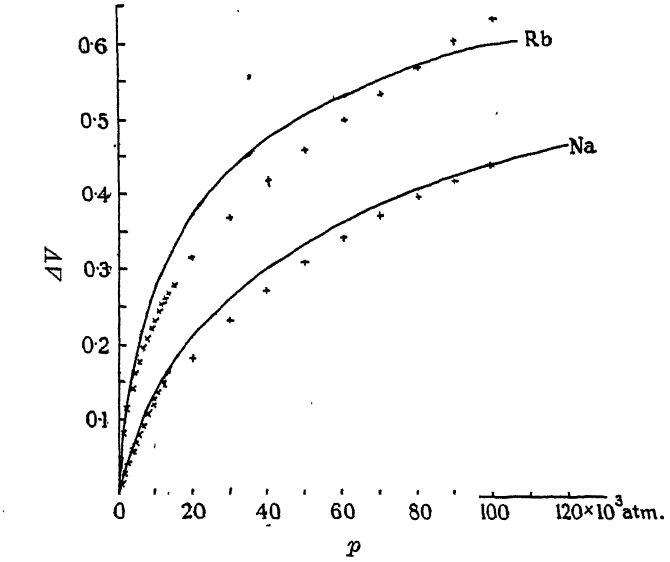


FIGURE 7. Compression curves of Na and Rb (× measured by Bridgman 1923, + measured by Bridgman 1941). $T = 300°$ K.

TABLE 4

element	$\delta\bar{V}/V_0$ (theor.)	$\delta\bar{V}/V_0$ (exp.)	element	$\delta\bar{V}/V_0$ (theor.)	$\delta\bar{V}/V_0$ (exp.)
Li	0·050	0·050	Tl ($T = 300°$)	0·024	0·026
Na	0·063	0·058	Sn	0·017	0·021
K	0·073	0·072	Sb ($T = 200°$)	0·014	0·005
Rb ($T = 300°$)	0·082	0·079	Sb ($T = 300°$)	0·025	0·008
Ca	0·028	0·025	Se ($T = 200°$)	0·033	0·031
Zn	0·033	0·026	Se ($T = 300°$)	0·050	0·046
Cd	0·030	0·027			

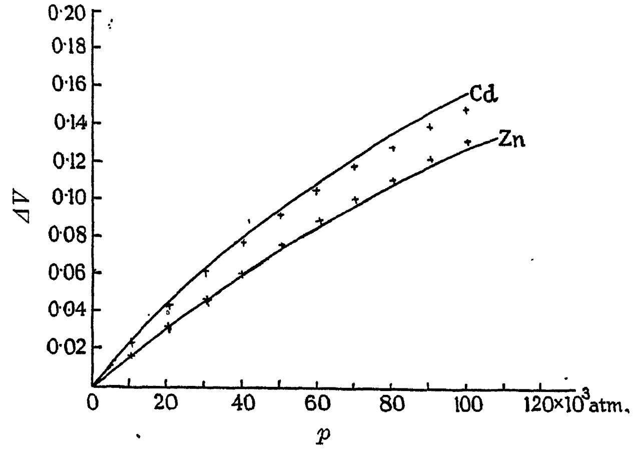


FIGURE 8. Compression curves of Cd and Zn (+measured by Bridgman 1941). $T = 300°$ K.

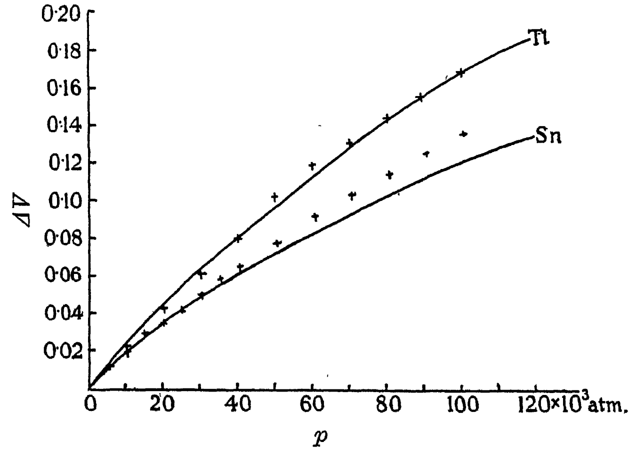


FIGURE 9. Compression curves of Tl and Sn (+measured by Bridgman 1938 and 1941). $T = 300°$ K.

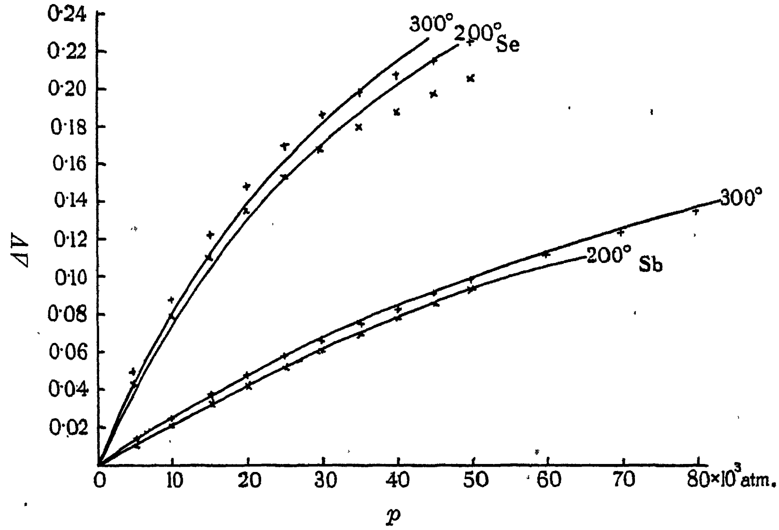


FIGURE 10. Compression curves of Se and Sb for different temperatures (+measured by Bridgman at room temperature 300° K, ×measured at CO$_2$ temperature 200° K, 1940).

melting-point and become infinitely large at this point itself. Observation shows that the actual increase of α with temperature is slight even up to T_m. Secondly, it has been shown by the present author in a former paper (Fürth 1941 *a*) that the value of $\delta V_c/V_0$, the relative increase of volume over its zero point value, calculated for the turning-point of the isobar $p = 0$ (according to the original theory of Born) is at least twice as big as the actually observed volume increase in a large number of cases.

It seems worth while to take up this question again on the basis of the new equation of state. The turning-point of the isobar $p = 0$ is determined by the condition

$$\left(\frac{\partial v}{\partial \tau}\right)_{\varpi = 0} = \infty, \tag{50}$$

which, in virtue of (6), can be written in the form

$$\left(\frac{\partial \tau}{\partial \xi}\right)_{\varpi = 0} = 0. \tag{51}$$

This leads, with the help of (5), (8) and (9), to the equation

$$(1+K\xi_c)\,[(b+c)+(b+Kc)\,\xi_c]\left(1+\frac{m+s}{s}\,\xi_c\right)-b\xi_c(1+\xi_c)\,(1-K) = 0 \tag{52}$$

for the value ξ_c of ξ corresponding to $T = T_c$ and $p = 0$. Using (11), (12) and (17) one can bring this equation into the form

$$C_0+C_1\xi_c+C_2\xi_c^2+C_3\xi_c^3 = 0, \tag{53}$$

where

$$\left.\begin{aligned} C_0 &= g,\\ C_1 &= g\left(2K+\frac{m+s}{s}\right),\\ C_2 &= K\left\{\left(1+\frac{m+s}{2}\right)\left(K+\frac{m}{s}\right)+\left(1+\frac{m+Ks}{2}\right)\left(2K+\frac{m}{s}\right)\right\},\\ C_3 &= \frac{s+m}{s}K^2\left(1+\frac{m+s}{2}\right). \end{aligned}\right\} \tag{54}$$

As ξ is always small compared with unity even up to ξ_c, a first approximation for solving the equation (53) can be tried by neglecting all terms except the first two. This leads to

$$\xi_c = -\left(2K+\frac{s+m}{s}\right)^{-1}. \tag{55}$$

In the same approximation one obtains from (5) (with $\varpi = 0$)

$$\tau_c = -\frac{a\xi_c}{b+c} = \frac{a}{g}\left(2K+\frac{s+m}{s}\right)^{-1}. \tag{56}$$

By means of (38) this expression can be transformed into

$$\frac{A}{T_c}=\frac{qR\theta}{T_c}=\frac{qR}{\tau_c}=\frac{qRg}{a}\left(2K+\frac{s+m}{s}\right), \tag{57}$$

and finally in consequence of (16)

$$\frac{A}{T_c}=\frac{3Rg}{m}\left(1+\frac{2Ks}{s+m}\right). \tag{58}$$

Similarly from (55) and (6)

$$\frac{V_c}{V_0}=v_c=(1+\xi_c)^{3/s}=\left[1-\frac{1}{2K+\dfrac{s+m}{s}}\right]^{3/s}. \tag{59}$$

As the right-hand sides of (58) and (59) can be calculated for any pair of values of m and n, the ratios A/T_c and V_c/V_0 can be evaluated. They are tabulated in the last two columns of table 1. In order to check these results the values of τ_c obtained from (56) were compared with those values of τ_c which can be graphically determined in the cases investigated in the paper of Gow. As in all these cases the agreement is perfect within the accuracy of the calculation, the approximation used can be regarded as justified. The values of v_c according to (59), however, are just about one-half of the actual values, which fact is, of course, due to the very steep ascent of the isobars near the turning-points. The figures in the last column of table 1 have therefore to be considered as fictitious, corresponding to an extension of the isobars in parabolic shape from low values of τ up to τ_c. The correct values could be easily obtained from (53) and (54) by numerical calculation, but this is hardly worth while, as τ_c is in all cases very much larger than τ_m, as will be shown immediately.

Table 5 contains in column 3 the values of A/T_m for the elements, calculated from the melting temperatures (column 2) and the sublimation energies (column 3, table 2). It also contains in column 4 the values of A/T_c determined from (58) by using the values of m and n previously obtained and tabulated in table 3. The ratio of these two quantities gives T_c/T_m which is tabulated in column 5. It is very nearly equal to one for the four rare gases; for all the other elements, with a few exceptions, it is roughly constant with an approximate value of 3·5.

It appears from this fact that T_m for all these substances is well inside the region of the isobars where they can be considered to be parabolic in shape. It seems reasonable therefore to compare the observed values of v_m with the fictitious values of v_c which follow from (59). This is also shown in table 5, which contains the observed values $\delta V_m/V_0$ in its column 6 (see Fürth 1941 *a*) and the values $\delta V_c/V_0$ in column 7. The ratio of these two quantities gives $\delta V_c/\delta V_m$, which is tabulated in the last column of table 5. It is equal to 1·6 for argon, the only rare gas for which data were available (Simon, Ruhemann & Edwards 1930), and it is again roughly constant for nearly all the other elements with an average of about 3·5.

The fact that the values of T_c/T_m are nearly the same as those of $\delta V_c/\delta V_m$ proves that the observed volume expansion at the melting-point is in agreement with the present theory. The fact that these values are roughly the same for most of the elements seems further to indicate that there might exist a relation between the mechanical breakdown of the lattice and fusion, but also that the two processes are certainly not identical. Melting occurs in a region where the lattice should be stable for all disturbances caused by the thermal movement. It is interesting to note that the ratio T_c/T_m for the rare gases is exactly equal to unity which can hardly be a mere coincidence. On the other hand, as $\delta V_c/\delta V_m$ is very much larger than unity in the case of argon it cannot be concluded that melting and mechanical instability of the lattice can be identified even in these cases.

TABLE 5

element	T_m	A/T_m (cal./deg.)	A/T_c (cal./deg.)	T_c/T_m	$(\delta V/V_0)_m$ $\times 10^3$	$(\delta V/V_0)_c$ $\times 10^3$	$\delta V_c/\delta V_m$
Ne	24·4	24·6	24·2	1·0	—	—	—
A	83·8	24·5	24·2	1·0	68	108	1·6
Kr	116	24·0	24·2	1·0	—	—	—
Xe	161	25·5	24·2	1·05	—	—	—
Li	453	79	(20)	3·9	(75)	310	4·1
Na	371	70	(18)	3·9	(72)	320	4·4
K	337	65	(18)	3·8	(81)	320	3·9
Rb	312	67	(18)	3·7	(82)	320	3·9
Cs	301	63	(18)	3·5	(81)	320	3·9
Cu	1356	60·5	(17)	3·5	69	240	3·5
Ag	1234	56	(17)	3·3	72	225	3·1
Au	1340	68	(18)	3·8	60	180	3·0
Be	1550	35	14·5	3·8	(78)	300	3·8
Mg	923	36·5	14·5	2·5	(75)	300	4·0
Ca	1080	40	14·5	2·8	(90)	300	3·3
Sr	1044	38	(14·5)	2·6	—	300	—
Ba	977	44·5	(14·5)	3·1	—	300	—
Zn	693	46	16	2·9	60	220	3·6
Cd	594	45·5	15	3·0	54	210	3·9
Hg	234	66	17	3·9	(30)	175	5·8
B	2500	(12)	(17)	0·7?	63	240	3·8
Al	932	72·5	17·3	4·2	66	240	3·6
Tl	576	74·5	22·5	3·3	(51)	170	3·3
C	3800	52	(14·5)	3·6	48	300	6·3?
Si	1693	55·5	(21)	2·6?	—	100	—
Sn	505	138	(28)	4·9	(36)	160	4·4
Pb	600	78	(34)	2·3	45	120	2·7
As	1087	25·1	(25)	1·1	(17)	220	13?
Sb	904	61	(17)	3·6	(25)	240	9·6
Bi	544	—	(17)	—	(22)	240	10·9
Cr	(2070)	(43)	16	2·7	(54)	220	4·1
Mo	2870	54	16	3·4	(54)	220	4·1
W	3650	55·5	16	3·4	72	220	3·1
Se	490	57	(12·5)	4·5	(75)	400	5·3
Mn	1493	46·5	(28)	1·7	(93)	150	1·6
Fe	1803	53	17	3·1	(69)	240	3·5
Co	1763	—	17	—	72	240	3·3
Ni	1725	56·5	17	3·3	78	240	3·1
Pt	2037	61·5	18	3·4	60	180	3·0

6

The rule about the melting-point referred to in the last paragraph can be expressed in the following form

$$T_m = \epsilon T_c, \tag{60}$$

where ϵ has roughly the same value for all substances. If it is assumed that this rule has a real physical meaning, then one can reasonably expect it to be closely connected with another well-known rule about melting, namely Lindemann's melting formula (Lindemann 1910). Such a connexion was first suggested by Born (1939), and it can be shown to exist as a consequence of the present theory.

The proper frequency ν_0 of the linear vibration of a pair of isolated atoms bound together by forces obeying the law (2) is given by

$$u = k\theta = \frac{4\pi^2}{mn}\mu\nu_0^2 r_0^2, \tag{61}$$

where μ is the mass of an atom. Neglecting the slight difference between r_0 and the actual distance between nearest neighbours in the lattice, one has further for the face-centred lattice

$$r_0 = \left(\frac{\sqrt{2}\,V_0}{N}\right)^{\frac{1}{3}} \tag{62}$$

(the numerical constant in the case of a body-centred lattice is practically identical). Introducing further the atomic weight M by

$$M = \mu N, \tag{63}$$

one obtains from (61) and (62) $$\nu_0^2 = \frac{\theta}{MV_0^{\frac{2}{3}}}\frac{RmnN^{\frac{2}{3}}}{2^{\frac{1}{3}}4\pi^2}. \tag{64}$$

Lindemann's formula contains a different atomic frequency, namely, the 'Einstein frequency' ν_E of an atom in the lattice, vibrating under the influence of all the other atoms. A simple consideration shows that when only first neighbours are taken into account and the forces from the more distant atoms neglected, the two frequencies ν_0 and ν_E are connected by $\nu_E^2 = 4\nu_0^2$ in a face-centred lattice, and by $\nu_E^2 = 8/3\nu_0^2$ in a body-centred lattice. In the first case the quantity q, defined by (14), is equal to 6, in the second case it is equal to 4; hence both the above formulae can be replaced by the single formula

$$\nu_E^2 = 2/3q\nu_0^2. \tag{65}$$

It will be supposed that this formula also holds approximately in the general case of the force law (2) with q given by (15).

Substituting from (64) into (65) and using the relations (3), (7), (16), (56) and (60), one obtains

$$\nu_E^2 = C_L^2\frac{T_m}{MV_0^{\frac{2}{3}}}, \tag{66}$$

where $$C_L = \sqrt{\left\{\frac{RgN^{\frac{2}{3}}}{2^{\frac{4}{3}}\pi^2\epsilon}[2K(n-m)+n]\right\}} = 1{\cdot}53\times 10^{11}\sqrt{\left\{\frac{g}{\epsilon}[2K(n-m)+n]\right\}}. \tag{67}$$

The relation (66) is indeed Lindemann's formula, where C_L should be a universal constant. If the atomic frequency ν_E is determined from the dependence of the specific heat on temperature, the formula can be checked, since all the other quantities contained in it are measurable. Such investigations have been carried out before (see e.g. Blom 1913), and they show that the law holds with considerable accuracy for a large number of elements, although C_L is by no means really strictly constant. Its average value is $3{\cdot}1 \times 10^{12}$. According to the present theory C_L should be given by formula (67). As its right-hand side depends on the exponents m and n it cannot be a constant which would account for the deviations from Lindemann's law. But the range of variation of C_L between the different elements is not very large and can be easily determined from the data of the tables 3 and 4. The calculation shows that C_L should be contained between $1{\cdot}8 \times 10^{12}$ and $3{\cdot}3 \times 10^{12}$. The experimental value quoted above is, indeed, within these limits.

REFERENCES

Austin, J. B. 1932 *Physics*, **3**, 240.
Austin, J. B. & Pierce, R. H. H. Jun. 1933 *Physics*, **4**, 409.
Blom, C. E. 1913 *Ann. Phys., Lpz.*, **42**, 1397.
Born, M. 1923 *Atomtheorie des festen Zustandes.* Leipzig.
Born, M. 1939 *J. Chem. Phys.* **7**, 591.
Born, M. 1940 *Proc. Camb. Phil. Soc.* **36**, 160.
Born, M. 1943 *Proc. Camb. Phil. Soc.* **39**, 100.
Born, M. & Bradburn, M. 1943 *Proc. Camb. Phil. Soc.* **39**, 104.
Born, M. & Landé, A. 1918 *Verh. dtsch. phys. Ges.* **20**, 210.
Bradburn, M. 1943 *Proc. Camb. Phil. Soc.* **39**, 113.
Bridgman, P. W. 1923 *Proc. Amer. Acad. Arts Sci.* **58**, 234.
Bridgman, P. W. 1925 *Proc. Amer. Acad. Arts Sci.* **60**, 385.
Bridgman, P. W. 1931 *The physics of high pressure.* London.
Bridgman, P. W. 1938 *Proc. Amer. Acad. Arts Sci.* **72**, 207.
Bridgman, P. W. 1940 *Proc. Amer. Acad. Arts Sci.* **74**, 21.
Bridgman, P. W. 1941 *Phys. Rev.* **60**, 351.
Debye, P. 1914 *Göttinger Vorträge über die kinetische Theorie der Materie*, pp. 17ff. Leipzig.
Eucken, A. & Dannöhl, W. 1934 *Z. Elektrochem.* **40**, 814.
Fürth, R. 1941*a* *Proc. Camb. Phil. Soc.* **37**, 34.
Fürth, R. 1941*b* *Proc. Camb. Phil. Soc.* **37**, 177.
Fürth, R. 1942 *Proc. Roy. Soc.* A, **180**, 285.
Gow, M. 1944 *Proc. Camb. Phil. Soc.* (in the Press).
Grüneisen, E. 1910 *Ann. Phys., Lpz.*, **33**, 1239.
Grüneisen, E. 1912 *Ann. Phys., Lpz.*, **39**, 257.
Grüneisen, E. 1926 *Zustand der festen Körper, Handb. d. Phys.* Berlin.
Herzfeld, K. & Goeppert-Mayer, M. 1934 *Phys. Rev.* **46**, 995.
Lennard-Jones, J. E. 1937 *Physica* (IV), **10**, 941.
Lindemann, F. A. 1910 *Phys. Z.* **11**, 609.
Mie, G. 1903 *Ann. Phys., Lpz.*, **11**, 657.
Misra, R. D. 1940 *Proc. Camb. Phil. Soc.* **36**, 175.
Mott, N. F. & Jones, H. 1936 *The theory of the properties of metals and alloys.* Oxford.
Nix, F. C. & MacNair, D. 1941 *Phys. Rev.* **60**, 597.
Nix, F. C. & MacNair, D. 1942 *Phys. Rev.* **61**, 74.
Roth, W. A. & Scheel, K. 1923–36 New edition of Landolt-Börnstein, *Physikalisch-chemische Tabellen.* Berlin.
Simon, F., Ruhemann, M. & Edwards, W. A. M. 1930 *Z. phys. Chem.* B, **6**, 331.
Slater, J. C. 1940 *Phys. Rev.* **57**, 744.
Stern, O. 1916 *Ann. Phys., Lpz.*, (4), **51**, 237.

Continuous spectra in flames: the role of atomic oxygen in combustion

By A. G. Gaydon, D.Sc., *Chemical Engineering Department, Imperial College, London, S.W. 7*

(*Communicated by Sir Alfred Egerton, Sec.R.S.—Received* 9 *November* 1943)

[Plate 2]

The causes and types of continuous spectra emitted by flames are discussed and their importance stressed. It is shown that the yellow-green continuous spectrum emitted by some flames containing oxides of nitrogen is probably identical with the spectrum of the air afterglow and is therefore due to a reaction between nitric oxide and atomic oxygen.

It thus becomes possible to test for the presence of atomic oxygen in a flame by admitting nitric oxide and observing if a yellow-green emission results. For the carbon monoxide flame there appears to be a high concentration of atomic oxygen, both for the dry and moist flame. The combustion mechanism is discussed in detail using this knowledge. For the hydrogen flame a little atomic oxygen is present, but results do not permit of definite conclusions. For hydrocarbon flames there is no sign of atomic oxygen in the inner cone, and this is taken as strong evidence in favour of a peroxidation rather than a hydroxylation mechanism.

Introduction

Of the three general classes of spectra observed in flames, line, band and continuous, the last have received least attention. Line spectra are, of course, due to electronic transitions between the quantized energy states of atoms; although of interest for the purpose of chemical analysis and in a few special cases such as the sodium-line reversal method of temperature measurement, line spectra are seldom of much importance in elucidating details of combustion mechanisms. Band spectra in the visible and ultra-violet regions correspond to similar transitions between quantized energy states of molecules; the emission and absorption of band spectra by flames have received a fair amount of attention in relation to combustion problems (for a general review see Gaydon 1942*a*). Continuous spectra (in gases) correspond to transitions in which at least one of the states involved is unquantized. Continua therefore correspond to *processes* such as dissociation, ionization, and association. Thus, while band spectra are of value in telling us what molecules and radicals are present under flame conditions, the examination of continuous spectra might be expected to give information of even greater value about the actual processes taking place during the combustion. Unfortunately, in practice, continuous spectra are difficult to identify with certainty owing to their characteristic lack of definite features. With band spectra, analysis of the structure gives precise information about certain molecular constants and so usually enables us to assign the bands to a particular molecule. With continuous spectra this is not possible, and in only a few cases, such as the well-known hydrogen continuum corresponding to the process $H_2(^3\Sigma_g^+) \rightarrow$ unstable $H_2(^3\Sigma_u^+) \rightarrow H + H$, can details of the process be stated with certainty.

When alkali metals, or their salts, or, to a less extent, alkaline earth metals are introduced into a flame, strong continuous emission around the blue or violet region is obtained. Hartley (1894) reported observations of this type of continuum which has also been observed by the author when potassium or rubidium are introduced into either a hydrogen or a carbon monoxide flame. The position of the maximum of this continuum depends on the ionization potential of the alkali metal; moving to longer wave-lengths for metals with lower ionization potential. This and other types of continuum have been very fully reviewed and discussed by Finkelnburg (1930, 1934, 1935) who has developed Stark's suggestion that the process responsible for this type of emission is the association of an electron and a positive ion to form a neutral molecule. Finkelnburg has used the term 'electron radiation' to describe this process and has shown that it is quantitatively of the right magnitude to account for the strength of the continuum.

All flames show strong ionization, and, even when there is no evidence to show that the main constituents (e.g. H_2 or O_2) are ionized, it is possible that small traces of impurity may result in enough ionization to give some continuous emission of the 'electron radiation' type as a background to the main features of the spectrum.

Continua are also emitted when halogens are introduced into flames. Urey & Bates (1929) have examined these and concluded that they result from the reunion of an excited halogen atom and a normal atom to form a normal halogen molecule. Several other inorganic flames, such as those of phosphorus, arsenic and selenium, also give strong continuous emission (for references see Gaydon 1942*a*), although their cause is unknown.

The spectrum of the bright blue carbon monoxide flame appears under small dispersion to be continuous. More detailed examination (Weston 1925; Gaydon 1940) reveals a very complex band structure. This is particularly well developed in the flame at reduced pressure and also in the cool flame (Gaydon 1943). In hot flames at high pressure the structure is less outstanding, and the appearance differs little from that of a pure continuum. There is no doubt that the higher temperature increases the length of the branches of the rotational fine structure of the band system and so contributes to this change of appearance. Whether or not there is a genuine continuous emission as well as the complex band spectrum is uncertain. The maximum temperature for the CO-O_2 flame is, at least in theory, relatively high, and appreciable ionization and dissociation, both factors favouring the emission of continua, will result from this high temperature.

The continuous emission from luminous organic flames, which is due to thermal emission from hot carbon particles, falls outside the scope of this investigation.

CONTINUOUS EMISSION BY THE HYDROGEN FLAME

The simple flame of hydrogen burning in air is practically non-luminous. The spectrum shows the OH bands in the ultra-violet and the vibration-rotation spectrum of H_2O in the near infra-red. The flame normally appears faintly orange, this coloration being due to traces of impurities, chiefly Na and CaO. Higher temperature flames of hydrogen do, however, show appreciable luminosity. Finkelnburg (1930) gives references to many old papers, some of which are not available to the author, as copies have been removed to safety for the duration of the war. Dibbits in 1864 found that hydrogen burning with air and other gases showed a weak continuum with maximum in the green. Frankland in 1868 observed that the spectrum of hydrogen burning in oxygen at ten atmospheres pressure was extremely bright and entirely continuous in the visible region, while Liveing & Dewar (1891) studied the rate of increase of the brightness of this continuum with pressure up to 40 atm.

The present experiments indicate that high-temperature flames of hydrogen burning with pure oxygen are blue, and that the pale yellowish green coloration only occurs in the presence of nitrogen. The blue luminosity is particularly well marked with a simple flame of oxygen burning in hydrogen at a quartz jet. This flame has a bright blue base, and is sometimes surrounded by a red outer mantle the colour of which is due to emission of the vibration-rotation bands of H_2O (Gaydon 1942*b*) which extend from the infra-red right up to the visible. A group of spectra of the blue part of this flame of oxygen burning in hydrogen is shown in the plate. There is strong emission throughout the blue, violet and near ultra-violet. The OH bands are very strong in this source and the very long branches of the rotational fine structure can be observed well down into the visible region, and there is some difficulty in distinguishing between true continuous emission and apparent continuum resulting from the unresolved fine structure of these OH bands. The blue colour is probably largely due to continuous emission, and this may be of the 'electron radiation' type. The possibility of association processes between free atoms present in these high-temperature flames cannot, however, be ruled out and may well contribute, along with other factors, to the blue colour.

The oxy-hydrogen blow-pipe flame burning in air has a blue core, but the outer part of the flame, especially the tip, shows a pale yellow or yellow-green luminosity. The colour is rather difficult to describe, and in this type of flame is usually confused by admixture with orange coloration due to the presence of Na and CaO as impurities; the true colour is probably best described as an unsaturated yellow-green, although it has on occasion been described as greyish green or brownish green or just as yellowish. It will be referred to as yellow-green throughout this paper. The same colour is produced when a trace of air or nitrogen is admitted to a flame of hydrogen burning in oxygen. Flames of hydrogen burning with nitrous oxide under suitable conditions show the yellow-green colour strongly. It was noticed that the oxy-hydrogen blow-pipe in air always smelt of oxides of nitrogen when the yellow-green coloration was well developed, and it seemed that there must be some connexion between the emission of this yellow-green continuum and some reaction involving an oxide of nitrogen. The colour is very similar to that of the afterglow of air in an electric discharge at reduced pressure. The author has therefore carried out a number of experiments to compare the yellow-green continuum obtained by deliberately introducing oxides of nitrogen into flames with that emitted by the air afterglow.

EXPERIMENTAL

The spectrograph employed was a large aperture (*f*/4) glass prism instrument. This gave a visible spectrum about 5 cm. long, this moderate dispersion in the visible region combined with high speed making the instrument particularly suitable for the study of the yellow-green continuum emitted by the relatively weak flame sources.

The photographic plate characteristics desired were uniform sensitivity throughout the visible region combined with fairly high contrast and speed. Several makes and varieties of plates were tried, but no really satisfactory type was found. All the panchromatic or 'long-range' types showed low or irregular sensitivity in the green, while orthochromatic types failed, of course, at the red end. Ilford Rapid Process Panchromatic were actually used for all the investigations, although these show a double maximum of sensitivity in the red and not very high sensitivity in the green, as may be seen in the group of spectra of a luminous Bunsen flame in plate 2*g*.

Refined measurements of the continuous spectra emitted by flames appear to be of little value owing to the variations introduced by temperature and other factors. It was found sufficient for the present purpose to take a series of five exposures (by racking down the plate holder) for each source and to compare these visually. Exposures in the ratio $1:\sqrt{3}:3:3\sqrt{3}:9$ were found suitable. A neon lamp comparison spectrum, showing lines of Ne, He and H, was used to give an approximate wave-length scale.

Nitric oxide was prepared by dropping 50 % sulphuric acid on to a concentrated solution of sodium nitrite and potassium iodide, the gas being stored over water and used undried except where otherwise stated (i.e. for CO flames). Other gases were commercial products.

The flames studied were of three types: flames of premixed gases burning at a quartz jet, flames of one gas burning in an atmosphere of another (for brief description of apparatus see Gaydon 1942*b*), and flames of premixed gases burning at a Smithell's separator. This separator consisted of two concentric quartz or pyrex tubes about 1 and 1·5 cm. bore, a cork or similar fitting enabling the smaller tube to be slid gently down until its top was about 5 cm. below that of the outer tube. By adjustment of gas flow and mixture strength it thus became possible to separate the inner cone of hydrocarbon flames on to the top of the smaller tube, leaving the outer cone burning at the top of the larger tube.

The simple discharge tube used for the examination of the air or oxygen afterglow is shown diagrammatically in figure 1. It consisted of two pyrex tubes A and B, each 30 cm. long and 1 cm. bore, with viewing windows of quartz, W_1, W_2 and W_3, and steel electrodes E_1 and E_2. Air or oxygen was admitted through a very fine capillary at the tap T_1 and was evacuated through tap T_3. Nitric oxide could be admitted at T_2 when desired. The discharge in the tube A was maintained by an induction coil. It was found that stray weak discharges to the pump were reduced by connecting the electrode E_2 and the pump so that they were at the same potential. The afterglow in the tube B was viewed end-on through the window W_2.

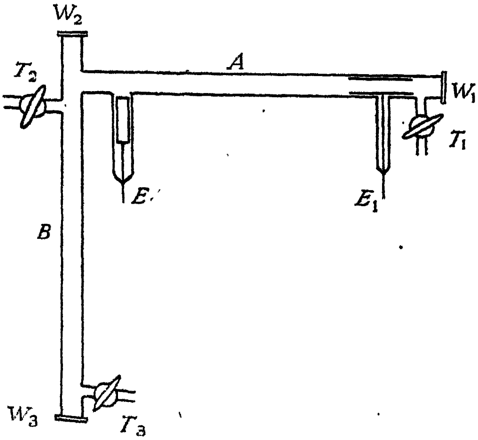


FIGURE 1. Discharge tube for afterglow.

THE YELLOW-GREEN CONTINUUM IN THE AFTERGLOW AND IN FLAMES

The yellow-green air, or so-called oxygen, afterglow, obtained when a discharge is passed through air or through oxygen containing a trace of nitrogen at a pressure of the order of a mm. of mercury, has been studied by a number of investigators (Rayleigh 1910; Stoddart 1934; Newman 1935), and Spealman & Rodebush (1935) have convincingly shown that the luminosity is due to a reaction between nitric oxide and atomic oxygen, this explanation fitting in well with the careful observations made earlier by Rayleigh. The spectrum is mainly continuous, extending throughout the visible region with a maximum somewhere in the green. Faint banded structure has also been observed, but it is very weak and features are difficult to measure.

The spectrum of the afterglow of air and that of oxygen containing a little nitrogen have been photographed under similar conditions to the flame spectra. It has also been found that, when a very slow stream of nitric oxide is admitted at tap T_2 (figure 1) and fairly pure oxygen is used in the discharge, a bright glow is obtained at the point where the gases mix. A group of spectra of this glow, due to the reaction of atomic oxygen formed in the discharge with the nitric oxide, is shown in plate 2*f*. The spectrum is continuous and of maximum intensity in the green; the double maxima in the red are due to plate sensitivity being high in these regions. There is some very faint banded structure, but even with the high contrast of the process plates this is not easy to see.

It has been remarked that some flames containing oxides of nitrogen show a yellow-green continuous emission. Plate 2*e* shows a group of spectra of the flame of nitrous oxide burning in hydrogen, which is of this type. It is also found that when nitric oxide is introduced into some, but not all, flames a strong yellow-green continuous emission results. This is true for carbon monoxide burning in oxygen (plate 2*d*) and, although rather less strongly, for hydrogen burning in oxygen (plate 2*c*).

The colour of the air afterglow and that of these flames containing nitric oxide is strikingly similar, and this is borne out by comparison of the spectra. The similarity of the colours of the afterglow to that obtained by introducing NO into a Bunsen flame was first pointed out by Rayleigh (1911). There seems no reason to doubt that the process responsible for the emission of these yellow-green continua is the same, namely, the reaction between nitric oxide and atomic oxygen. The cause of the yellow-green colour of flames containing oxides of nitrogen has thus been ascertained.

The absorption spectrum of the molecule NO_2 shows a banded structure in the visible region, indicating that absorption in this region results in the formation of an electronically excited stable NO_2 molecule. Beyond about 3700 A the band structure merges into a continuum, indicating the process

$$NO_2 + h\nu = NO + O.$$

There is thus the possibility that a collision between nitric oxide and atomic oxygen will result in the reverse process, the emission of a light quantum and the formation of a normal NO_2 molecule. This process would result, in the absence of appreciable thermal energy in the colliding particles, in a continuous emission with short-wave limit around 3700 A. Assuming an electron transition probability of the order 10^7 or 10^8 per sec. and a duration of the order 10^{-12} or 10^{-13} sec. for a collision, we obtain a probability of about the order 10^{-5} for the reaction

$$NO + O = NO_2 + h\nu$$

to occur as a bimolecular association. This seems to be of the right magnitude. In the afterglow each oxygen atom will make about 10^7 collisions per sec., and, as only a small proportion of these collisions will be with nitric oxide molecules, this gives a reasonable duration for the afterglow, which in practice is usually of the order of a second, although much less if nitric oxide is admitted.

There is also a possibility of a triple collision between nitric oxide, an oxygen atom and a third body. In this case the third body might take up the rather large amount of energy set free by the formation of NO_2, or alternatively it might take up only a small part of this energy leaving the excess as electronic energy of excitation of the newly formed NO_2 molecule. In this latter case the NO_2 molecule might then emit a banded spectrum. This would account for the faint banded emission superposed on the yellow-green continuum of the afterglow spectrum.

The majority of ordinary flames in air do not contain appreciable quantities of nitric oxide so that even if atomic oxygen is present the yellow-green continuum will not be emitted. For the very hot oxy-hydrogen flame, however, small quantities of nitric oxide may be formed if air is allowed to mix in with the flame, and so the flame shows the faintly luminous tip which has been commented on. Similarly, nitrous oxide usually breaks up into oxygen and molecular nitrogen rather than to oxides of nitrogen, and so the yellow-green continuum is not necessarily emitted even if atomic oxygen is present; under certain flame conditions, such as nitrous oxide burning in hydrogen, however, it seems that both nitric oxide and atomic oxygen are present, and so the flame shows the continuous emission; all flames of this type smell strongly of oxides of nitrogen.

NITRIC OXIDE IN FLAMES

The assignment of the yellow-green continuum to a reaction between nitric oxide and atomic oxygen gives us a simple qualitative test for the presence of atomic oxygen in a flame by introducing a little nitric oxide and observing whether or not the yellow-green continuous emission is then obtained. In this section the experimental results obtained by subjecting a number of flames to this test are described; application of the results to a discussion of the combustion mechanism is deferred to later sections.

Carbon monoxide. Flames, of all types, of carbon monoxide burning with air or oxygen are bright blue in colour. In all cases the addition of a little nitric oxide produces a marked change to a whitish yellow-green. The effect is particularly strong when the NO is added to a flame of CO burning in O_2. Groups of spectra illustrating this are shown in plate 2*a* and *d*. It may be noted that the addition of a little nitrous oxide instead of nitric oxide does not produce the same effect.

The original experiments on CO flames were carried out with undried gases. In view of the importance of moisture in the combustion of CO, a flame of $CO + NO$ in O_2 in which the gases were carefully dried was also examined. The CO and O_2 were dried by slow passage through a liquid oxygen trap, and the NO was passed first through sulphuric acid and then through three U tubes packed with phosphorus pentoxide. The yellow-green colour was again produced with about the same intensity. These experiments cannot be compared with those on the combustion of intensively dried carbon monoxide, but they serve to show that moderate drying does not materially affect the concentration of atomic oxygen in the flame, which remains high.

Flames of carbon monoxide burning with nitrous oxide are bluish in colour and show a rather similar spectrum to that of flames with oxygen, but the blue shade is somewhat less intense, and this has already been commented on (Gaydon 1942*b*). This departure from the bright blue colour is particularly noticeable for the premixed CO-N_2O flame, and indicates some slight nitric oxide formation, the reaction of which with atomic oxygen leads to yellow-green emission which dulls the brightness of the blue colour of the flame. The addition of a very little nitric oxide to these flames supported by nitrous oxide produces a very marked intensification of the yellow-green colour.

Hydrogen. When a little nitric oxide is added to a flame of hydrogen burning in oxygen a definite, although not strong, coloration results. The colour is rather greyish green, the departure from the yellow-green shade being apparently the result of admixture with the natural blue shade of the untreated flame. With the flame of oxygen burning in hydrogen, which is quite a bright blue, the effect of adding nitric oxide is less marked. When the NO is admitted in the H_2 stream there does not appear to be any appreciable effect on the colour of the flame, which, however, enlarges slightly. When the NO is admitted in the O_2 stream there does appear to be a slight greenish coloration, but not very marked. In either case, however, there is a definite effect on the outer flame of the excess hydrogen burning in air, which becomes grey-green and shows the continuous emission spectrum. Groups of spectra of hydrogen burning in oxygen and the same with a little nitric oxide added are shown in plate 2*b* and *c*.

Flames of hydrogen burning with nitrous oxide are of rather varied type (Dixon & Higgins 1926). In some cases the spectrum shows NO, NH and the ammonia α bands (Fowler & Badami 1931) and in other cases, such as N_2O burning in H_2 or the premixed flame containing excess of N_2O, the yellow-green continuum is strong and dominates the spectrum.

Methane. Flames of methane in oxygen or air or vice versa are luminous (carbon deposition) and are not suitable for the nitric oxide test. The premixed CH_4-air flame, however, gives striking results. With a simple premixed flame of the Bunsen type it was observed that fairly strong yellowish green coloration was produced by the addition of nitric oxide, and that this colour was apparently restricted to the outer zone of the flame. This was confirmed by study of the separated cones in a Smithell's separator.

The methane-air flame in the Smithell's separator shows a bright blue-green inner cone and a very pale blue-violet outer cone. On admitting nitric oxide the inner cone turns bright violet and the outer cone shows the yellow-green colour strongly. This is a very simple and effective experiment.

The spectrum of the inner cone, before the admission of nitric oxide, shows the usual C_2 Swan and CH bands. When nitric oxide is present the CN Violet system comes up strongly, this accounting for the violet colour, and the C_2 Swan bands are partly suppressed; the CH bands do not appear to be affected (see plate 2*h*). With a steadily burning flame there is not the slightest trace of the yellow-green continuous emission in the inner cone. Only with an unsteady flame, or with excess air, does a slight evanescent yellow-green colour appear above and around the inner of the separated cones, and this is clearly due to incomplete separation of the two types of combustion occurring in the two cones.

Ethylene. The results with ethylene are similar to those with methane, but the yellow-green coloration of the outer cone is rather less strong unless a fairly large amount of nitric oxide is present. This may be due to the partial decomposition of the nitric oxide in this very hot flame.

Butane. The results for butane are also similar to those for methane. In the Smithell's separator the inner cone changed from blue-green to violet when nitric oxide was added, while the outer cone changed from pale blue-violet to bright yellow-green.

THE COMBUSTION OF CARBON MONOXIDE; THE ROLE OF ATOMIC OXYGEN

The experimental results detailed above serve as a qualitative test for the presence of atomic oxygen in the various flames studied. For the carbon monoxide flame it is clear that a considerable concentration of atomic oxygen exists, both for the dry and moist gases. Previously the author (Gaydon 1942*a*) has been sceptical of combustion mechanisms involving atomic oxygen, but it now seems that this attitude must be modified in discussing the carbon monoxide flame as there can be little doubt that the high concentration of atomic oxygen is the result of the combustion processes and is not merely thermal in origin. It should be remembered that the concentration of atomic oxygen will depend on the processes consuming the atoms as well as on those responsible for their liberation.

The several proposed combustion mechanisms for moist and dry CO have been dealt with in books by Hinshelwood (1933), Semenoff (1935) and Lewis & von Elbe (1938). For the dry reaction it is very difficult to devise a mechanism involving only CO, O_2, O and CO_2 which will account satisfactorily for the chemical kinetics. Lewis & von Elbe have therefore supposed that ozone and a hypothetical carbon peroxide CO_3 may take part in the reactions. Of the reactions included as likely by them, the following lead to the formation of atomic oxygen:

$$O_3+CO = CO_2+O+O-15\,\text{kcal.}, \qquad (1)$$

$$O_3+CO = CO_3+O, \qquad (2)$$

and the following to its removal:

$$O+O_2+M = O_3+M, \qquad (3)$$

$$O+CO+M = CO_2+M. \qquad (4)$$

They rightly stress that at room temperature reaction (4) is at least 500 times slower than (3) and is therefore probably unimportant as a chain-breaking process.

The flame spectrum of CO shows a banded spectrum on a continuous background. This banded spectrum is believed (Gaydon 1940) to be emitted by CO_2. Kondratjewa & Kondratjew (1937) have shown that under favourable conditions at least one quantum of light is emitted by the flame for every 125 normal CO_2 molecules formed. This is vastly too much radiation to result from thermal causes, and in any case there is no evidence that this spectrum can be obtained from CO_2 molecules by either thermal or electrical excitation. Of the processes retained by Lewis & von Elbe, (4) is the only one liberating sufficient energy for the formation of electronically activated CO_2 molecules

$$CO + O + M = CO_2^* + M. \tag{4a}$$

This process is probably slightly endothermic. It might be followed by

$$CO_2^* = CO_2 + h\nu \quad (\textit{banded}), \tag{4b}$$

by

$$CO_2^* + O_2 = CO_2 + O + O, \tag{4c}$$

which is probably slightly exothermic, or by reactions involving the dissociation of the CO_2^* or another CO_2 molecule.

It has already been stressed by the author that there is no evidence that the electronically excited CO_2 molecules are metastable. Their life is probably between 10^{-8} and 10^{-5} sec. This would, however, allow for a fair number of collisions with oxygen molecules and, although sufficient radiation by process (4*b*) must take place to give the observed band spectrum, the majority of the excited molecules will probably end in reaction (4*c*). Thus reaction (4) should be regarded as *chain-branching* and not chain-breaking.

It is usual, in discussing combustion mechanisms, to disregard direct bimolecular associations because of the difficulty of removing the excess energy from the system. For the special case of the carbon monoxide flame, however, it is clear that the bimolecular association

$$O + NO = NO_2 + h\nu,$$

resulting in the yellow-green continuous emission, is able to compete for atomic oxygen with the combustion processes, and therefore there seems no reason why the similar association

$$O + CO = CO_2 + h\nu \quad (\textit{continuous}) \tag{5}$$

may not also have some probability. The reaction (5) probably requires some activation energy for a favourable collision accompanied by radiation, and the probability will in any case be less than 10^{-5} at each collision, but nevertheless the process may account for some of the continuous background to the flame spectrum. The existence of an activation energy for this reaction fits in with the observation that the continuous background is weaker, relative to the banded spectrum, for flames at low pressure and temperature and especially for the cool flame (Gaydon 1943).

It has been stressed (Gaydon 1942*a*) that normal carbon dioxide, CO_2 ($X^1\Sigma$), cannot be formed from normal CO ($X^1\Sigma$) and oxygen, O (^3P), without electronic rearrangement. Thus reaction (1) probably requires a very high activation energy in addition to the 15 kcal. absorbed in the reaction, and may therefore be of little importance. These considerations also apply to reaction (4) but not to (4*a*). Little can be said about the hypothetical reaction (2). From a spectroscopic viewpoint it seems that reactions (4*a*), (4*b*), (4*c*) and (5) are required to account for the spectrum and the presence of atomic oxygen in the dry flame.

It should, however, be stressed that perfectly dry CO will not readily burn at all (Ubbelohde 1933), and that although the reactions listed above may serve to maintain combustion they do not account for its initiation. It is likely that for reaction (4*a*) to take place the third body M may have to fulfil special conditions and that water is particularly suitable.

For the moist reaction the concentration of atomic oxygen in the flame is also high. It seems likely, from the present study, that there is no essential difference in the combustion mechanism in the presence of a trace of water, but that its thermal decomposition may serve to initiate the reaction chains, and that by collisions the water molecules may quickly set free the vibrational energy of the newly formed CO_2 molecules so that this energy can thermally assist the further reaction instead of merely producing abnormal dissociation of the CO_2 in the flame products. However, the reaction chain

$$OH+CO = CO_2+H+27\,\text{kcal.}, \quad H+O_2 = OH+O-18\,\text{kcal.},$$

which could be initiated by thermal decomposition of H_2O to OH and H, would also lead to accumulation of atomic oxygen and may be important.

The hydrogen flame

For flames in which oxygen is in excess, i.e. for hydrogen burning in oxygen or air, there is a moderate concentration of atomic oxygen, although not as high as that in the CO flame. When hydrogen is in excess, i.e. oxygen burning in hydrogen, the concentration of atomic oxygen appears to be small.

The results do not appear to be sufficiently definite to draw any certain conclusions. It seems that some atomic oxygen may be formed by the reaction

$$H+O_2 = OH+O-18\,\text{kcal.}, \tag{6}$$

and that the removal of this atomic oxygen in the presence of an atmosphere of hydrogen may be more rapid than in an atmosphere of oxygen, but in the absence of quantitative data it cannot be said whether this reaction is of major importance in the combustion mechanism for hydrogen.

The mechanism of hydrocarbon combustion

The experiments on methane, ethylene and butane in the Smithell's separator show clearly that in the outer cone, corresponding to the later stages of the combustion when hydrogen and carbon monoxide are the principal combustibles, there is a high concentration of atomic oxygen, but that in the inner cone, corresponding to the early stages of the combustion of the hydrocarbons, atomic oxygen is not present in sufficient quantity to give the slightest coloration with the nitric oxide test. It is known that for methane at room temperature there is practically no reaction with atomic oxygen and that methane does not show an 'atomic flame' with oxygen. It therefore seems fairly certain that the negligible concentration of atomic oxygen in the early stages of the combustion must be due to its not being formed rather than to its very rapid removal. This receives experimental support from the observation that a flame of mixed carbon monoxide and butane (with a fairly high proportion of CO and air) burning in a Smithell's separator does show the yellow-green coloration in the inner cone, indicating presumably that the butane is not particularly effective in removing the atomic oxygen.

Two main types of mechanism, and a number of variants on these, have been suggested to explain the initial stages of hydrocarbon oxidation. These are usually referred to as the hydroxylation and peroxidation theories. The hydroxylation theory, put forward by the late Professor W. A. Bone, assumes the successive introduction of hydroxyl groups into the hydrocarbon molecule, and its most promising modification, that proposed by Norrish, makes use, for methane, of the chain process

$$O+CH_4 = CH_4O = CH_2+H_2O, \quad CH_2+O_2 = H_2CO+O.$$

According to the peroxidation theory as suggested by Engler and by Bach and applied to combustion processes by Callendar and by Sir Alfred Egerton and others, reaction occurs when a sufficiently energetic oxygen molecule combines momentarily with an energetic hydrocarbon molecule to form a temporary peroxide which subsequently breaks down to aldehyde and water, setting free energy to maintain the combustion. Variants on this theory proposed by Ubbelohde and by von Elbe & Lewis make use of reaction chains involving, for methane, the radical CH_3 and either another peroxidic radical or OH. These combustion mechanisms have been discussed elsewhere (Gaydon 1942*a*), and it has been shown that certain spectroscopic observations give some support for a peroxidation rather than a hydroxylation theory.

These new experiments, showing a negligible amount of atomic oxygen in the inner cone, appear to provide evidence against Norrish's modification of the hydroxylation theory which requires atomic oxygen for the chain reactions.

It is a pleasure to express my thanks to Sir Alfred Egerton for his close interest in the work and helpful discussions. I am also indebted to the Council of the Royal Society for generous financial assistance.

References

Dixon, H. B. & Higgins, W. F. 1926 *Mem. Manchr Lit. Phil. Soc.* **71**, 15.
Finkelnburg, W. 1930 *Phys. Z.* **31**, 1.
Finkelnburg, W. 1934 *Phys. Rev.* **45**, 341; **46**, 330.
Finkelnburg, W. 1935 *Z. Phys.* **93**, 201.
Fowler, A. & Badami, J. S. 1931 *Proc. Roy. Soc.* A, **133**, 325.
Gaydon, A. G. 1940 *Proc. Roy. Soc.* A, **176**, 505.
Gaydon, A. G. 1942*a* *Spectroscopy and combustion theory.* London: Chapman and Hall.
Gaydon, A. G. 1942*b* *Proc. Roy. Soc.* A, **181**, 197.
Gaydon, A. G. 1943 *Proc. Roy. Soc.* A, **182**, 199.

Hartley, W. N. 1894 *Phil. Trans.* A, **185**, 161.
Hinshelwood, C. N. 1933 *Kinetics of chemical change in gaseous systems.* Oxford.
Kondratjewa, E. & Kondratjew, V. 1937 *J. Phys. Chem. Russ.* **9**, 736.
Lewis, B. & von Elbe, G. 1938 *Combustion flames and explosions of gases.* Cambridge.
Liveing, G. D. & Dewar, J. 1891 *Proc. Roy. Soc.* A, **49**, 217.
Newman, F. H. 1935 *Phil. Mag.* **20**, 777.
Rayleigh, Lord (R. J. Strutt) 1910 *Proc. Phys. Soc.* **23**, 66.
Rayleigh, Lord (R. J. Strutt) 1911 *Proc Roy. Soc.* A, **86**, 56.
Semenoff, N. 1935 *Chemical kinetics and chain reactions.* Oxford.
Spealman, M. L. & Rodebush, W. H. 1935 *J. Amer. Chem. Soc.* **57**, 1474.
Stoddart, E. M. 1934 *Proc. Roy. Soc.* A, **147**, 454.
Ubbelohde, A. R. 1933 *J. Chem. Soc.* p. 972.
Urey, H. C. & Bates, J. R. 1929 *Phys. Rev.* **34**, 1541.
Weston, F. R. 1925 *Proc. Roy. Soc.* A, **109**, 177, 523.

For description of plate see page 124.

DESCRIPTION OF PLATE

Groups (*a*)–(*g*) each show at the top the spectrum of a neon glow lamp for approximate wavelength scale, and below five spectra of the various sources studied with exposures increasing in the ratio $1:\sqrt{3}:3:3\sqrt{3}:9$. Ilford Process Panchromatic plates were used, and conditions of development, etc., were maintained as constant as possible throughout.

(*a*) Flame of CO burning in O_2. Exposures 10–90 sec.

(*b*) Flame of O_2 burning in H_2. Exposures 2–18 min.

(*c*) Flame of H_2 mixed with NO burning in O_2. Exposures $\frac{1}{2}$–$4\frac{1}{2}$ min.

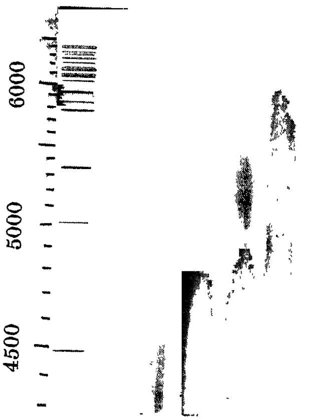
(*d*) Flame of CO mixed with NO burning in O_2. Exposures 10–90 sec.

(*e*) Flame of N_2O burning in H_2. Exposures 1–9 min.

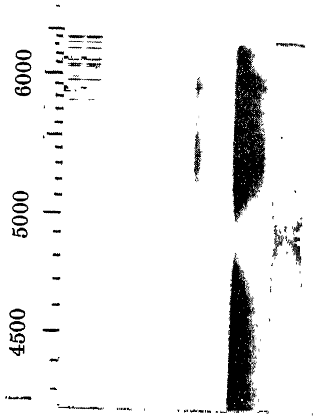
(*f*) Glow produced by mixing nitric oxide with afterglowing oxygen from electric discharge. Exposures $6\frac{2}{3}$–60 min.

(*g*) Luminous Bunsen flame. Exposures 5–45 sec. This serves to illustrate photographic plate sensitivity.

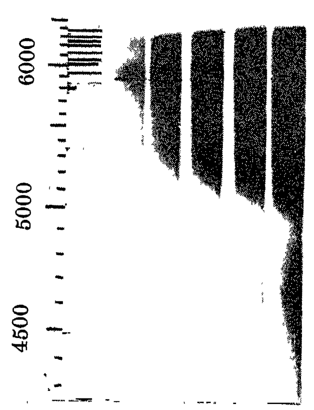
(*h*) Upper spectrum; inner cone of flame of premixed methane and air in Smithell's separator. Exposure 10 min. Lower spectrum; the same with nitric oxide added.



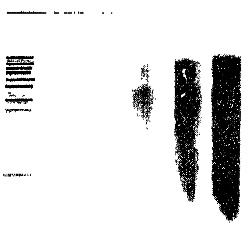
(a) CO on O_2



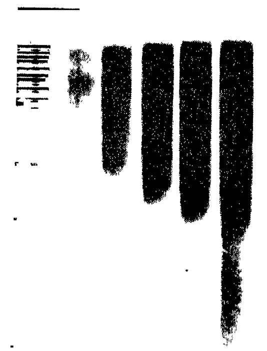
(b) O_2 in H_2



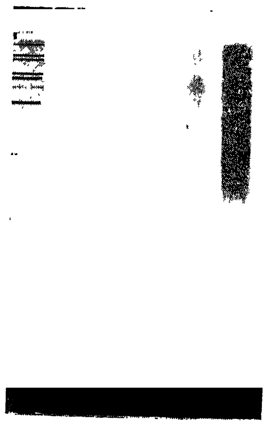
(c) H_2+NO in O_2



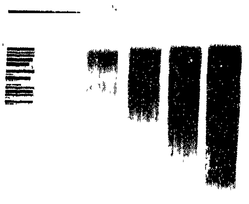
(d) CO+NO in O_2



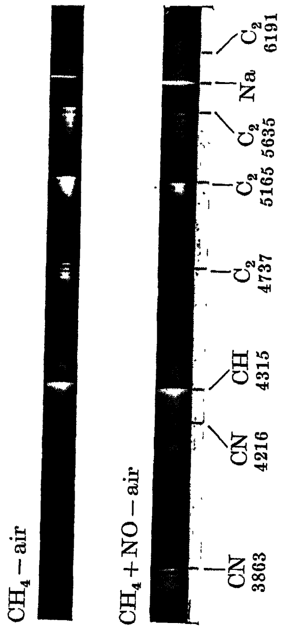
(e) N_2O in H_2



(f) NO in O_2 afterglow



(g) luminous Bunsen flame



(h) inner cones

Relaxation methods applied to engineering problems
X. The graphical representation of stress

By D. N. de G. Allen, B.A. and R. V. Southwell, F.R.S.

(*Received* 27 *January* 1944)

A convenient representation of a vector quantity (e.g. the shear stress on a specified plane) is by contours which exhibit its resultant magnitude and by 'trajectories' which have its direction at every point. But it sometimes happens that only its components in two perpendicular directions are known, and then, while the contours are easy to construct, derivation of the 'trajectories' presents a special problem. If this can be solved, tensor quantities also will be representable, e.g. a state of plane stress.

By orthodox methods the problem would appear to be difficult, but in this paper it is shown to yield to 'relaxational' attack.

Introduction

1. This paper deals with a class of problem having importance in various branches of mathematical physics, and special importance in the theory of elasticity. Its worked examples all relate to the representation of stress, and its title has been chosen accordingly; but it has much wider application.

Graphical representation of a vector quantity

2. A vector quantity has magnitude and direction, therefore requires two quantities for its specification. For example, the shear stress on a plane (x, y) can be specified by its components

$$X_z = S\cos\theta, \quad Y_z = S\sin\theta, \tag{1}$$

S denoting the resultant shear intensity, θ the inclination of its direction to the axis Ox. If now it is desired to present this specification graphically, an obvious means of doing so will be (*a*) a diagram showing by contours the variation of S, (*b*) a family of 'stress trajectories' which, being everywhere inclined at θ to Ox, have the direction of the resultant shear.

From (1) we have

$$S^2 = X_z^2 + Y_z^2, \tag{2}$$

whence S can be calculated when X_z, Y_z are known (e.g. from experimental measurements); so construction of the contours (*a*) presents no difficulty. To construct the stress trajectories (*b*), the simplest procedure is to regard them as contours of some function β not yet determined. Then

$$\cos\theta\frac{\partial\beta}{\partial x} + \sin\theta\frac{\partial\beta}{\partial y} = 0, \tag{3}$$

since β has zero gradient in the direction θ; and combining (3) with (1) we see that (provided S is nowhere infinite)

$$X_z\frac{\partial\beta}{\partial x} + Y_z\frac{\partial\beta}{\partial y} = 0. \tag{4}$$

Our problem is to deduce β from this equation, which implies that

$$X_z = -\alpha\frac{\partial\beta}{\partial y}, \quad Y_z = \alpha\frac{\partial\beta}{\partial x}, \tag{5}$$

α being some other function of x and y, not yet determined.

3. Usually, in the theory of elasticity, though X_z and Y_z are not known initially, expressions of the type of (5) can be assigned to them.* Our problem is the converse of this: X_z, Y_z being given, the corresponding distributions of α and β must be found in order that β-contours (the 'stress trajectories') may be constructed. The function α will be easy to deduce when β has been found, since

$$\alpha^2\left\{\left(\frac{\partial\beta}{\partial x}\right)^2 + \left(\frac{\partial\beta}{\partial y}\right)^2\right\} = X_z^2 + Y_z^2 = S^2, \tag{6}$$

by (5) and (2).

GRAPHICAL REPRESENTATION OF A TENSOR QUANTITY

4. In general stress is a tensor, not a vector quantity, having three components even in the two-dimensional case. In a state of 'plane stress', for example,

$$Z_x = Z_y = Z_z = 0 \tag{7}$$

and the other stress components have the expressions

$$X_x = \frac{\partial^2\chi}{\partial y^2}, \quad Y_y = \frac{\partial^2\chi}{\partial x^2}, \quad X_y = -\frac{\partial^2\chi}{\partial x\,\partial y} \tag{8}$$

in terms of a stress function χ which satisfies the 'biharmonic equation'

$$\nabla^4\chi = 0. \tag{9}$$

More generally, i.e. when (7) are not satisfied, three further quantities are entailed. But Z_z can be exhibited by contours, and the shear stress of which Z_x, Z_y are components can be represented by contours and by 'stress trajectories' in the manner of §§ 2–3; so the only problem remaining is to represent, for all points in a specified plane (x, y), a state of stress specified by its components X_x, Y_y, X_y when these are *not* given by functional expressions of the type of (8).

---

* Thus the stresses induced by torsion in a rod of circular section but non-uniform diameter have the expressions

$$\widehat{\theta z} = -\frac{1}{r^2}\frac{\partial\phi}{\partial r}, \quad \widehat{r\theta} = \frac{1}{r^2}\frac{\partial\phi}{\partial z} \tag{i}$$

in terms of a function ϕ which has a constant value on the generator of the stress-free boundary, and which satisfies the equation

$$\frac{\partial}{\partial r}\left(r^{-3}\frac{\partial\phi}{\partial r}\right) + \frac{\partial}{\partial z}\left(r^{-3}\frac{\partial\phi}{\partial z}\right) = 0$$

at every point in the material. These expressions become identical with (5) if, in the latter, z, r, θ, r^{-2}, ϕ are substituted for x, y, z, α, β.

Love (1927, § 57) cites a mode of presentation suggested by Michell (1901), but implies that a complete representation of stress is not practicable. It would seem, however, to present no special difficulty once the problem of vector representation (§§ 2–3) has been solved; for a stress defined by X_x, Y_y, X_y will be completely represented by (*a*) a set of orthogonally intersecting trajectories directed along the planes of principal shear stress, together with (*b*) contours exhibiting the local intensity of the principal shear stress and (*c*) contours exhibiting the distribution of (X_x+Y_y), a quantity which is invariant for change of axes. The intensity (*b*) of the principal shear stress is given by

$$|S| = \tfrac{1}{2}\sqrt{\{(X_x-Y_y)^2+4X_y^2\}}, \tag{10}$$

and (*a*) the planes of principal shear stress are inclined to Ox at angles θ_1, θ_2 which are roots of the equation

$$\cot 2\theta = -2X_y/(X_x-Y_y). \tag{11}$$

Thus all of $|S|$, θ_1, θ_2, (X_x+Y_y) are calculable, and (*b*) and (*c*) present no difficulty. To plot the trajectories (*a*) we require to know the distribution of some function β such that

$$\cos\theta\frac{\partial\beta}{\partial x}+\sin\theta\frac{\partial\beta}{\partial y} = 0, \tag*{(3) bis}$$

when θ is identified either with θ_1 or θ_2. *This problem has been considered already.*

OUTLINE OF THE RELAXATIONAL ATTACK

5. Graphical representation of stress, *whether this be a vector or a tensor quantity*, thus entails the finding of a function β which satisfies an equation of the type of

$$X_z\frac{\partial\beta}{\partial x}+Y_z\frac{\partial\beta}{\partial y} = 0, \tag*{(4) bis}$$

in which X_z, Y_z are known. This is a purely mathematical problem, not yet attacked by Relaxation Methods. For orthodox (analytical) methods it would seem to be in general difficult, even when X_z, Y_z are specified functions of x and y; and insoluble when, as will often happen, their functional forms are not known.*

It is clear that solutions will be in some degree arbitrary, since (4) is still satisfied when β is replaced by any function of β. For the same reason, any values may be given to β along contours which are known initially, either as being parts of a stress-free boundary, or from considerations of symmetry. But there is nothing arbitrary in the *shapes* of the resulting contours, which alone have practical importance; and the problem yields without difficulty to a relaxational treatment, $\partial\beta/\partial x$ and $\partial\beta/\partial y$

---

* In hooks, for example, the shear stress on a cross-section can be related (along with the other four stress-components) with two stress functions ϕ and ψ (Southwell 1942, § 26); but the stress functions have not so far been determined analytically, and although they can be found by Relaxation Methods, these give values of the stress-components only at nodal points of a chosen net.

in (4) being replaced, as usual, by finite-difference approximations. Details will be best explained in relation to particular examples.

Example 1

6. It can be shown (Love 1927, § 231 (*f*)) that, when Poisson's ratio $\sigma = \frac{1}{4}$,

$$X_z = x^2 - a^2, \quad Y_z = \tfrac{1}{2}xy \tag{12}$$

represent (with neglect of a constant factor) the shear stress called into play when a uniform shearing action is applied to a beam of which the axis is Oz and cross-sections are bounded by the curve

$$x^2/a^2 + 16y^4/a^4 = 1. \tag{13}$$

This curve is symmetrical with respect both to Ox and to Oy, and clearly Ox is one of the wanted contours, since at every point on it $Y_z = 0$. Accordingly β may be given the values zero along Ox, 100 along the curved boundary (13). Since β must also have symmetry about Oy, computation may be confined to a single quadrant.

This problem has been treated with results which are shown in figure 1: fine lines give the correct (theoretical) contours, bold lines the contours derived by Relaxation Methods on the fairly coarse net shown. The computed values (shown to the left of nodal points) are not sufficiently numerous to determine the forms of the stress trajectories near A; but here their exact shapes do not matter, since the shear stress tends to zero at A.*

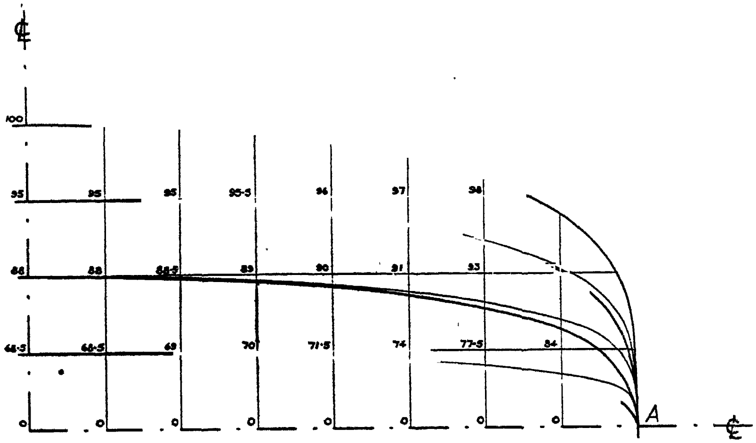


Figure 1

7. The finite-difference approximation to (4) is

$$\mathbf{F}_0 = (X_z)_0(\beta_1 - \beta_3) + (Y_z)_0(\beta_2 - \beta_4) = 0, \tag{14}$$

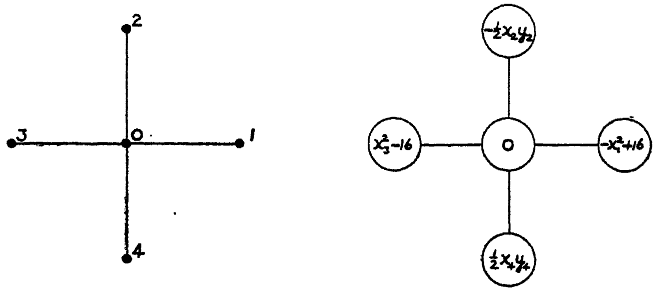
i.e. in this instance,

$$\mathbf{F}_0 = (x_0^2 - 16)(\beta_1 - \beta_3) + \tfrac{1}{2}x_0 y_0(\beta_2 - \beta_4) = 0, \tag{15}$$

---

* Because 'shear stress cannot cross an unloaded boundary'. Cf (e.g.) Southwell 1941, § 129.

the suffixes 0, 1, 2, 3, 4 relating to points so denoted in figure 2*a*. When 0 lies on Ox, (15) is satisfied identically, and when 0 lies on Oy it is satisfied in virtue of symmetry (in both instances because $\beta_1 = \beta_3$). From (15) we deduce the 'relaxation pattern' shown in figure 2*b*, which represents the effect on the F's of a unit increment $\Delta\beta_0$ made to the value of β at an isolated point 0. The use of such 'patterns' has been explained in previous papers of this series: it is in no way altered (although the labour of 'liquidation' is increased) by the fact that here a different pattern holds for every point. 'Irregular stars' can be treated in the manner of Part III (Christopherson & Southwell 1938, §§ 23–24).



(a) (b)

FIGURE 2

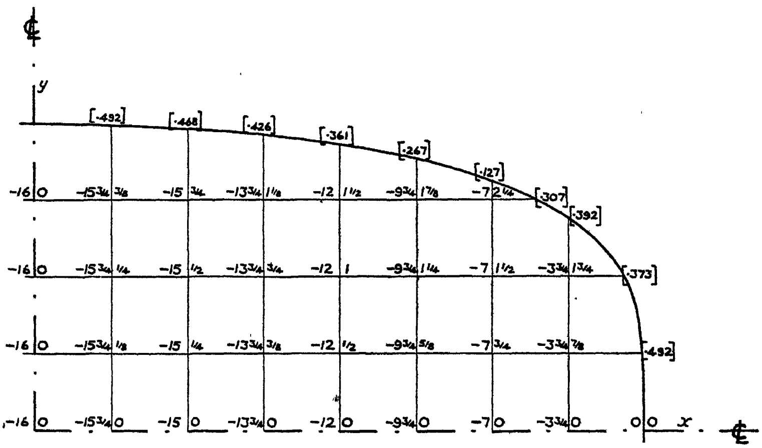
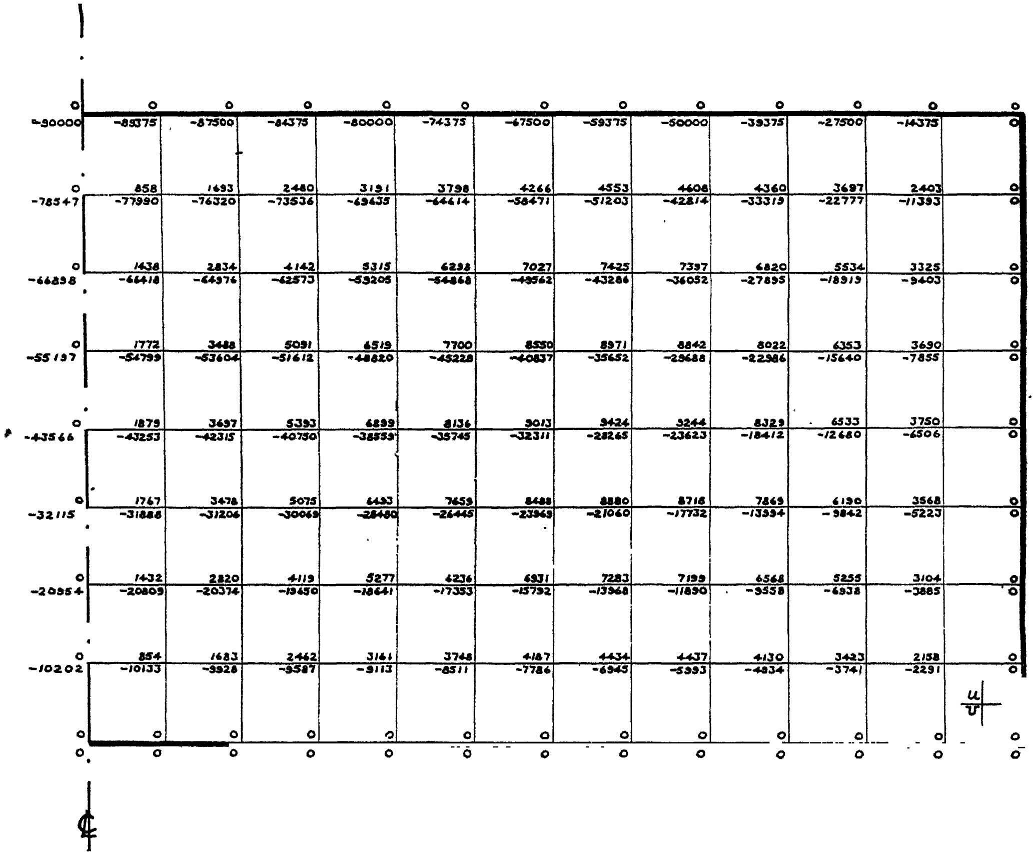


FIGURE 3

In figure 3, where the semi-axes are 4 and 2, so that $a = 4$ in (12) and (13), values of $(x^2 - 16)$ and of $\frac{1}{2}xy$ are recorded at nodal points, and the lengths of all 'irregular' strings are shown (in square brackets). Trial values for β gave calculable 'residuals' at the internal nodes: these were liquidated by driving them towards the axis Oy, where β, being unrestricted, could be altered to liquidate residuals at neighbouring (internal) nodes.

Example 2

8. Our second example is of the kind discussed in § 4. We seek a graphical representation of the solution found in an earlier paper of this series to a problem in 'plane strain',—namely, the effect of specified boundary displacements* imposed on a rectangle having sides in the proportion 3 : 1.



0 -90000	0 -89375	0 -87500	0 -84375	0 -80000	0 -74375	0 -67500	0 -59375	0 -50000	0 -39375	0 -27500	0 -14375	0 0
0 -78547	858 -77990	1693 -76320	2480 -73536	3191 -69435	3798 -64614	4266 -58471	4553 -51203	4608 -42814	4360 -33319	3697 -22777	2403 -11393	0 0
0 -66898	1438 -66418	2834 -64976	4142 -62573	5315 -59205	6298 -54868	7027 -49562	7425 -43286	7397 -36052	6820 -27895	5534 -18919	3325 -9403	0 0
0 -55197	1772 -54799	3488 -53604	5091 -51612	6519 -48820	7700 -45228	8550 -40837	8971 -35652	8842 -29688	8022 -22986	6353 -15640	3690 -7855	0 0
0 -43566	1879 -43253	3697 -42315	5393 -40750	6899 -38559	8136 -35745	9013 -32311	9424 -28265	9244 -23623	8329 -18412	6533 -12680	3750 -6506	0 0
0 -32115	1767 -31888	3478 -31206	5075 -30069	6493 -28480	7659 -26445	8488 -23969	8880 -21060	8716 -17732	7869 -13994	6190 -9842	3568 -5223	0 0
0 -20954	1432 -20809	2820 -20374	4119 -19650	5277 -18641	6236 -17353	6931 -15792	7283 -13968	7199 -11890	6568 -9558	5255 -6938	3104 -3885	0 0
0 -10202	854 -10133	1683 -9928	2462 -9587	3161 -9113	3748 -8511	4187 -7786	4434 -6945	4437 -5993	4130 -4934	3423 -3741	2158 -2291	0 0
0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0

FIGURE 4

A diagram in the earlier paper presented its solution in the form of displacement values computed for nodal points of a square-mesh net: for this investigation those values were recalculated to one more significant figure (figure 4), and from them distributions of $(X_x - Y_y)$ and of X_y were derived and recorded in figure 5. (For present purposes absolute magnitudes are not material.) It would of course be possible to deduce the corresponding stress function χ: the problem here, however, is to replace figure 5 by diagrams of the type of (*a*), (*b*) and (*c*), § 4, *without* having recourse to explicit functional expression.

---

* As figure 4 shows, $u = 0$ on all four edges, and $v = 0$ on all except the top edge, where it has a parabolic distribution tending to induce compression.

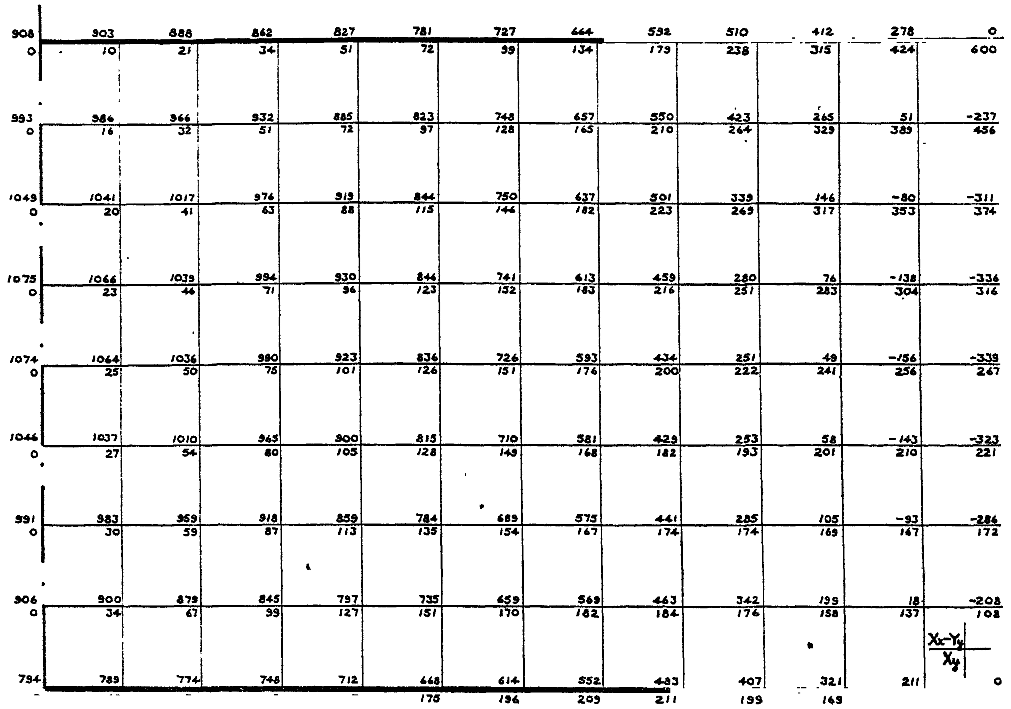


FIGURE 5

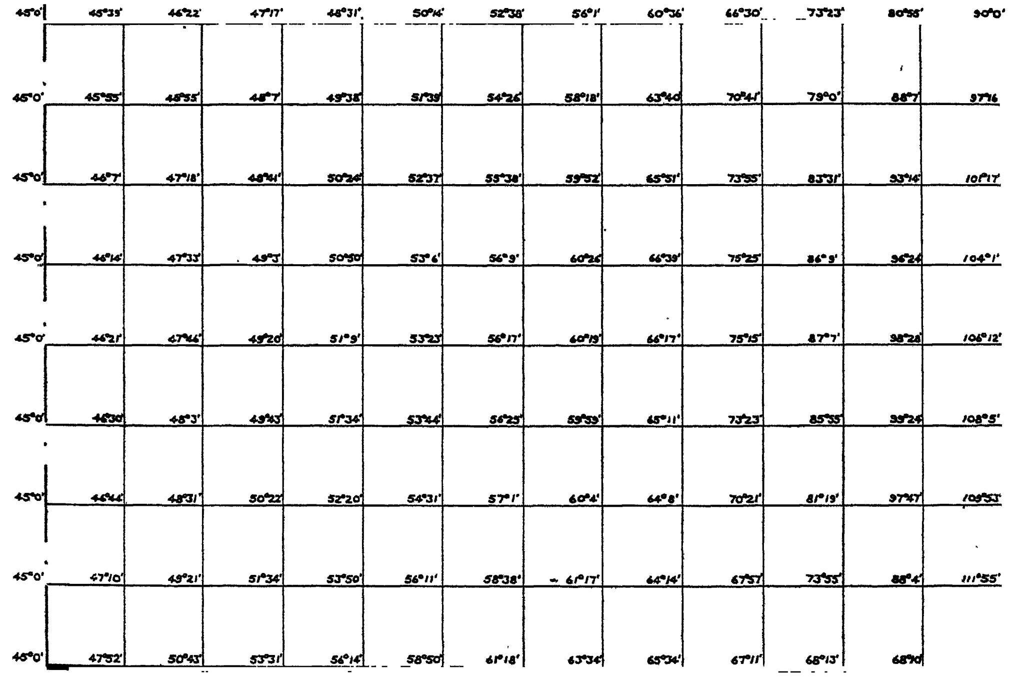


FIGURE 6

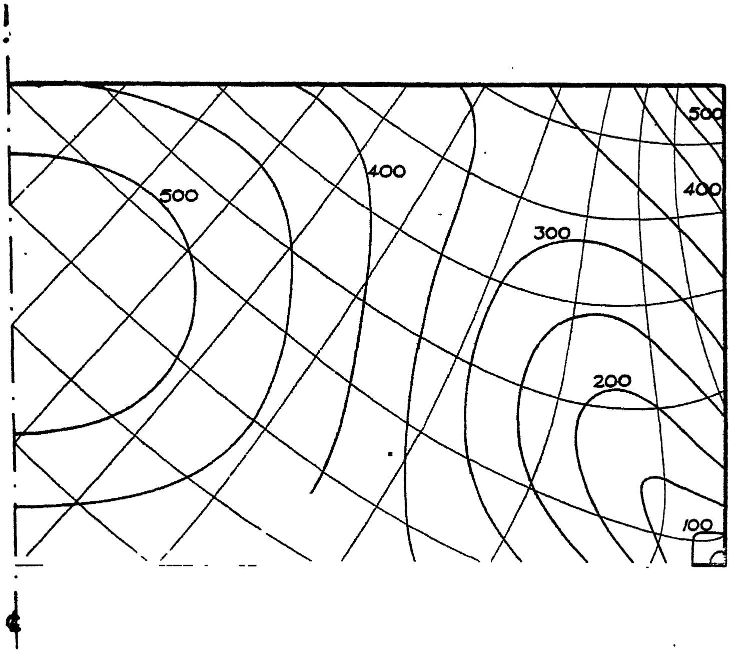


FIGURE 7. Example 2: trajectories and contours of principal shear stress.

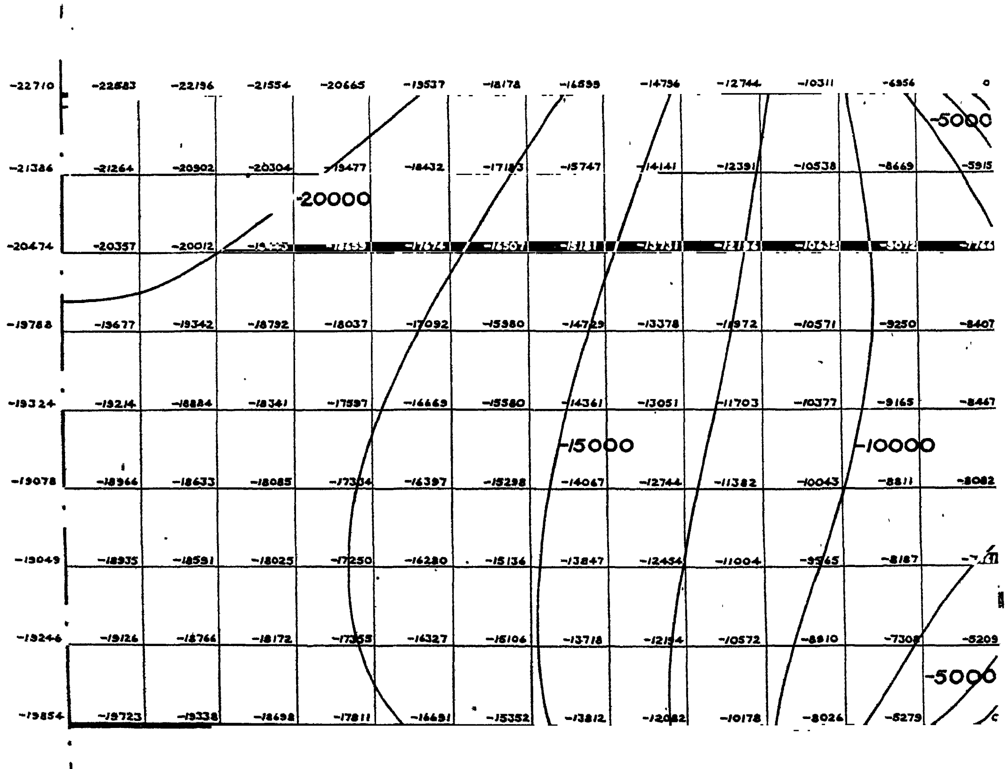


FIGURE 8. Example 2: contours of the stress sum $(X_x + Y_y)$.

9. Figure 6 records the nodal values of θ_1, one of the two roots of equation (11): the second root $\theta_2 = \theta_1 + 90°$. There is no question of the value appropriate to each family of 'trajectories', and for each family computation proceeds on the lines of § 7, but from (3) as governing equation instead of (4). No feature calls for special notice.

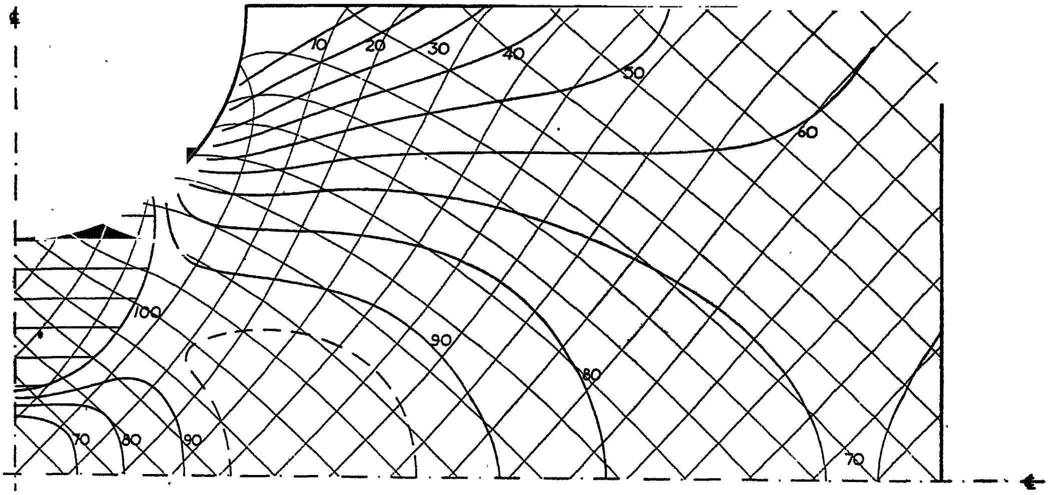


FIGURE 9. Example 3: trajectories and contours of principal stress difference (= 2 × principal shear stress).

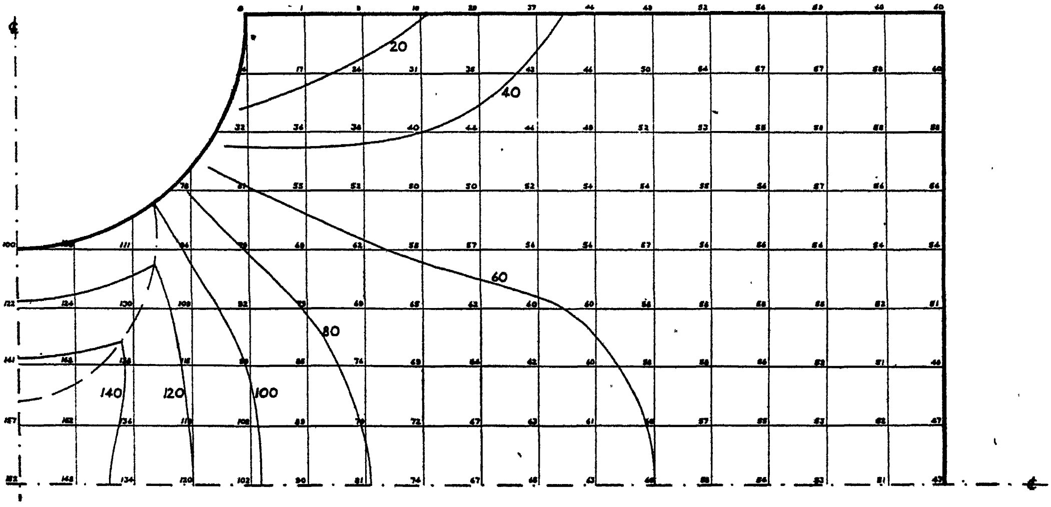


FIGURE 10. Example 3: contours of the stress sum $(X_x + Y_y)$.

The two orthogonally intersecting families are exhibited in figure 7, together with bold-line contours, deduced from (10), which give the intensity of the principal shear. Figure 8 shows the variation of the stress sum $(X_x + Y_y)$ as deduced from figure 4. Here too, no novel problem is confronted.

Example 3

10. Figures 9 and 10 (corresponding with figures 7 and 8) show the same problem solved for a stress system which is partly plastic, partly elastic, so that *the contours exhibit 'refraction' at a plastic-elastic interface.* The stress system is that induced by tension applied along the horizontal centre-line of a symmetrical specimen having two semicircular notches which form a 'waist'. The computed stress-components, being based on a particular hypothesis regarding plastic strain, are open to question; but this is of no importance to the present paper, which is concerned solely with the graphical representation of a *specified* stress.

REFERENCES

Christopherson, D. G. & Southwell, R. V. 1938 ('Part III') *Proc. Roy. Soc.* A, **168**, 317–350.
Love, A. E. H. 1927 *The mathematical theory of elasticity*, 4th ed. Camb. Univ. Press.
Michell, J. H. 1901 *Proc. Lond. Math. Soc.* **32**, 35–61.
Southwell, R. V. 1941 *Introduction to the theory of elasticity*, 2nd ed. Oxford Univ. Press.
Southwell, R. V. 1942 *Proc. Roy. Soc.* A, **180**, 367–96.

On the theory of point-particles

BY H. J. BHABHA, F.R.S. AND HARISH-CHANDRA

Cosmic Ray Research Unit, Indian Institute of Science, Bangalore

(*Received* 20 *March* 1944)

It is deduced from the conservation of the energy-momentum tensor that if the flow of energy and momentum into a tube surrounding a time-like world-line, on which the field is singular, become singular as the size of the tube is contracted to zero, then the singular terms are necessarily perfect differentials of quantities on the world-line with respect to the proper time along the world-line. The same can be proved of any other tensor, as, for example, the angular-momentum tensor, which is conserved. It is proved from this that for any *point*-particle whatever having charge, spin or other properties, which need not be specified, it is always possible to deduce exact equations of motion which are finite.

It is proved further that if the energy-momentum tensor is altered by the addition of $\partial K^{\mu\nu\sigma}/\partial x^{\sigma}$, where $K^{\mu\nu\sigma}$ is any tensor antisymmetric in ν and σ, then the equations of motion are unaltered, but it is possible to choose $K^{\mu\nu\sigma}$ in such a way as to make the flow of energy and momentum into a given tube non-singular.

By a point-particle is understood a particle whose field-producing and inertial properties are all located at a point. The particle may have a finite charge and a finite mass, but the charge density and mass density are exactly zero at every point of space other than the point at which the particle is located at that instant of time. The motion of the particle through space-time is therefore described by a time-like world-line. If the particle possesses a dipole or a higher multipole moment, then this is described by a suitable co-ordinate having a given value at each point of the world-line, but it is again assumed that the dipole density is exactly zero at all points of space-time not lying on the world-line. Thus the motion of the particle through space-time is completely described by a time-like world-line with the values of the co-ordinates describing the spin and other properties of the particle given at each point of it.

It has been shown by Dirac (1938) that an exact theory of a point electron moving in an electromagnetic field can be set up free from singularities, and Bhabha (1940) and Bhabha & Corben (1941) have shown that a similar theory free from singularities can be set up for a point dipole. It has been shown further by one of us (Bhabha 1939, 1941) that the theory can be extended to point-particles interacting with meson fields. These cover all the cases of practical interest, but a general demonstration that it is always possible to set up an exact classical theory for point-particles interacting with any field has not so far been given. This will be done in the present paper.

In all the work referred to above, the method used for finding the equations of motion of the point-particle is the same. From the field equations we calculate the field produced by the point-particle, or more exactly, we take that solution of the *homogeneous* field equations which has a singularity of the required type on the world-line. We now surround a finite length of the world-line by a world tube whose radius is ultimately made to tend to zero, and calculate the flow of energy, momentum and angular momentum into the world tube from outside by using the usual energy-momentum and angular-momentum tensors of the field. For brevity we shall refer to the quantities so calculated as the inflow. The equations of motion are now found from the condition that conservation of energy, momentum and angular momentum require that the flow of all these quantities into the tube must only depend on conditions at the two ends of the tube, that is, it must only be a function of the co-ordinates of the particle and their higher derivatives and also possibly of the field quantities at the two ends of the tube. Since the field is singular on the world-line, the usual energy-momentum tensor is also singular, and in consequence the flow of energy and momentum into the tube is likewise singular. That in spite of this it has always been possible to derive finite equations for the motion of the point-particles has always depended on the circumstance that the singular parts of the inflow over an infinitesimal length of the tube are always perfect differentials. We shall prove that this is a general property which is a consequence only of the conservation of the energy-momentum tensor and therefore that finite equations for the motion of a point-particle of any type whatsoever can always be derived.

We shall also show that it is always possible to modify the energy-momentum tensor by the addition of the divergence of a tensor of higher rank in the manner suggested by Pryce (1938) so that the inflow over a tube of given shape becomes finite. But the inflow would not necessarily be finite over a tube of any other shape. That it is always possible to modify the tensor in the same way as to make finite the inflow over tubes of any arbitrary shape, as also the energy and momentum integrals over arbitrary space-like surfaces is shown by one of us (H.-C.) in the paper which immediately follows this.

1. We use the same notation as in the previous papers by one of us (H.J.B.). We take the metric tensor to have the form $g_{00} = -g_{11} = -g_{22} = -g_{33} = 1$ with $g_{\mu\nu} = 0$ for $\mu \neq \nu$. The world-line of the particle is described by four co-ordinates $z^\mu(\tau)$ which are functions of the proper time τ measured from some point along it. The other co-ordinates of the particle, if any, need not be specified for the present work. A dot denotes differentiation with respect to τ, and $V^\mu = \dot{z}^\mu$ is used to denote the velocity of the particle. x^μ denotes a point of space, $s^\mu \equiv x^\mu - z^\mu(\tau_0)$ the distance from any point of space to the retarded point τ_0 on the world-line defined by

$$s_\mu s^\mu = 0. \tag{1}$$

If τ_0 be kept fixed then equation (1) is also the equation of the light cone whose apex is at τ_0. We further introduce a quantity κ as a function of the co-ordinates x^μ defined by

$$\kappa \equiv s_\mu v^\mu(\tau_0). \tag{2}$$

The energy-momentum tensor is denoted by $T^{\mu\nu}$. It satisfies the conservation equation

$$\frac{\partial T^{\mu\nu}}{\partial x^\nu} = 0. \tag{3}$$

It is necessary to find the generalization of Gauss's theorem to four-dimensional hyperbolic space. Let

$$\zeta^0(x^\mu) = \alpha \tag{4}$$

be a closed three-dimensional surface S surrounding a four-dimensional volume V such that the surface $\zeta^0(x^\mu) = \alpha - \Delta\alpha$ for positive $\Delta\alpha$ is contained inside (4) and lies wholly in the volume V. Let ζ^1, ζ^2 and ζ^3 be parameters defining the position of a point on the surface S. Then, if X^μ be any tensor, the generalization of Gauss's theorem reads

$$\int_V \frac{\partial X^\nu}{\partial x^\nu} dx^0 dx^1 dx^2 dx^3 = \int_S X^\nu \frac{\partial \zeta^0}{\partial x^\nu} |D|^{-1} d\zeta^1 d\zeta^2 d\zeta^3, \tag{5}$$

where D is the determinant of the transformation from the x's to the ζ's and the surface is covered in such a way that $d\zeta^1$, $d\zeta^2$, $d\zeta^3$ are always positive:

$$D \equiv \begin{vmatrix} \frac{\partial \zeta^0}{\partial x^0} & \frac{\partial \zeta^1}{\partial x^0} & \frac{\partial \zeta^2}{\partial x^0} & \frac{\partial \zeta^3}{\partial x^0} \\ \frac{\partial \zeta^0}{\partial x^1} & \frac{\partial \zeta^1}{\partial x^1} & \frac{\partial \zeta^2}{\partial x^1} & \frac{\partial \zeta^3}{\partial x^1} \\ \frac{\partial \zeta^0}{\partial x^2} & \frac{\partial \zeta^1}{\partial x^2} & \frac{\partial \zeta^2}{\partial x^2} & \frac{\partial \zeta^3}{\partial x^2} \\ \frac{\partial \zeta^0}{\partial x^3} & \frac{\partial \zeta^1}{\partial x^3} & \frac{\partial \zeta^2}{\partial x^3} & \frac{\partial \zeta^3}{\partial x^3} \end{vmatrix}. \tag{6}$$

$\partial\zeta^0/\partial x^\mu$ is the normal to the surface S. For a displacement $(\partial\zeta^0/\partial x^\mu)\Delta l$ in the direction of this normal with positive Δl the change $\Delta\zeta^0$ of ζ^0 is

$$\Delta\zeta^0 = \frac{\partial\zeta^0}{\partial x^\mu}\frac{\partial\zeta^0}{\partial x_\mu}\Delta l.$$

This is positive if $\partial\zeta^0/\partial x^\mu$ is a time-like vector and negative if it is space-like. Thus on the space-like portions of the surface the normal to the surface on the right-hand side of (5) is directed *outwards*, and on the time-like portions it is directed *inwards*. By writing $dS_\nu = \dfrac{\partial\zeta^0}{\partial x^\nu}\,|\,D\,|^{-1}\,d\zeta^1 d\zeta^2 d\zeta^3$ for an element of the surface S with its normal directed in the sense defined above, the right-hand side of (5) can be written

$$\int X^\nu dS_\nu. \tag{7}$$

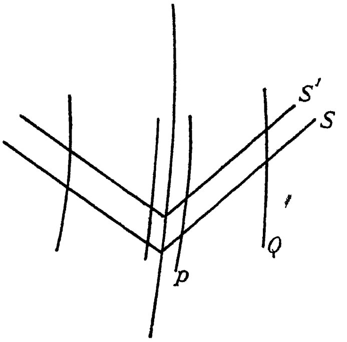


FIGURE 1

Consider a portion of the world-line between the points τ_1 and τ_2 and surround it by a world-tube of any arbitrary shape. We consider the ends of the tube to be given by the two two-dimensional surfaces where it intersects the future light cones from the points τ_1 and τ_2 respectively. The flow of energy and momentum into the tube is then given by

$$\int_{\tau_1}^{\tau_2} T^{\mu\nu} dS_\nu, \tag{8}$$

the normal to the surface being directed outwards. For conservation, (8) must be put equal to a function of the form $A^\mu(\tau_2) - A^\mu(\tau_1)$, where $A^\mu(\tau)$ is a function of v^μ and the other co-ordinates of the particle, possibly also of the field quantities, and their higher derivatives at the point τ only. On differentiating this equation

$$\frac{d}{d\tau}\int_{\tau_1}^{\tau} T^{\mu\nu} dS_\nu = \dot{A}^\mu(\tau). \tag{9}$$

This is not a mathematical identity but furnishes the equations of motion of the point-particle. It should be noted at once that the equations derived in this way are independent of the shape of the tube. For if the given tube is surrounded by another of some other shape, then it follows from (3) and (5) that the difference in the integral (8) over the two tubes is just equal to the integral over the portions of the light cones at τ_1 and τ_2 intercepted between the two tubes. In the limit when the size of both tubes is made to tend to zero, the integrals over the light cones just become functions of the conditions at τ_1 and τ_2 respectively, and hence the difference in (8) calculated over two tubes of different shape is identically of the form $A'^{\mu}(\tau_2)-A'^{\mu}(\tau_1)$. This difference can be absorbed into the right-hand side of (9), and hence it does not affect the equation of motion which can be so derived. One may therefore work with whatever tube is most convenient.

Consider now a volume V lying between the two light cones S and S' starting from the points τ' and $\tau'+d\tau'$ respectively, and between the two tubes P and Q defined by $\kappa=\epsilon$ and $\kappa=\eta$, as shown in figure 1. The boundary of V so defined can be characterized by the equation $\zeta^0=0$, where ζ^0 is a discontinuous function chosen in the following way:

$$\zeta^0=\begin{cases}\tau-(\tau'+d\tau') & \text{on } S',\\ \kappa-\eta & \text{on } Q,\\ \epsilon-\kappa & \text{on } P,\\ \tau'-\tau & \text{on } S.\end{cases}$$

Since (3) holds inside this volume V it follows from (5) that

$$\int_{\tau=\tau'+d\tau'} T^{\mu\nu}dS_\nu-\int_{\tau=\tau'} T^{\mu\nu}dS_\nu=\int_{\kappa=\epsilon} T^{\mu\nu}dS_\nu-\int_{\kappa=\eta} T^{\mu\nu}dS_\nu, \tag{10}$$

the normals to the light cones being taken in the direction of increasing τ in both cases, and towards the world-line on the two world-tubes. To evaluate the second integral on the left it is convenient to take $\zeta^1=\kappa$, $\zeta^2=s^2$ and $\zeta^3=s^3$ in addition to $\zeta^0=\tau$. As shown in the previous papers, it follows from (1) and (2) that

$$\frac{\partial\tau}{\partial x^\mu}=\frac{s_\mu}{\kappa}, \tag{11}$$

$$\frac{\partial\kappa}{\partial x^\mu}\equiv\kappa_\mu=v_\mu-\frac{s_\mu}{\kappa}(1-\kappa'),\quad \text{where}\quad \kappa'\equiv\dot{v}_\mu s^\mu, \tag{12}$$

$$\frac{\partial s^2}{\partial x^\mu}=\delta^2_\mu-v^2\frac{s_\mu}{\kappa}, \tag{13}$$

$$\frac{\partial s^3}{\partial x^\mu}=\delta^3_\mu-v^3\frac{s_\mu}{\kappa}. \tag{14}$$

The right-hand sides of (11)–(14) have to be inserted as the columns of the matrix on the right of (6). Since the determinant of a matrix vanishes if it has any two columns identical, it follows that we may omit all except the first terms on the right-hand sides of (11)–(14) and get

$$D=\frac{s_0v_1-s_1v_0}{\kappa}. \tag{15}$$

Therefore

$$\int_{\tau=\text{const.}} T^{\mu\nu}dS_\nu = \iiint_\epsilon^\eta T^{\mu\nu}\frac{s_\nu}{\kappa}\left|\frac{\kappa}{s_0v_1-s_1v_0}\right| ds^2\,ds^3\,d\kappa = \iint_\epsilon^\eta T^{\mu\nu}\frac{s_\nu}{\kappa}\,d\Omega\,d\kappa, \tag{16}$$

$d\Omega$ being the element of surface of a sphere of radius κ in the 'rest system' in which $v_0 = 1$, $v_1 = v_2 = v_3 = 0$. Similarly,

$$\int_{\kappa=\epsilon} T^{\mu\nu}dS_\nu = \int d\tau \int_{\kappa=\epsilon} T^{\mu\nu}\kappa_\nu\,d\Omega. \tag{17}$$

(10) now becomes
$$\frac{d}{d\tau}\left\{\int_\epsilon^\eta \int_\kappa T^{\mu\nu}\frac{s_\nu}{\kappa}\,d\Omega\,d\kappa\right\} = \int_{\kappa=\epsilon} T^{\mu\nu}\kappa_\nu\,d\Omega - \int_{\kappa=\eta} T^{\mu\nu}\kappa_\nu\,d\Omega. \tag{18}$$

It is important to note that the left-hand side of (18) is *not* a perfect differential. The integral does not depend in general only on conditions on the world-line at the point τ. Differentiating this equation with respect to ϵ, we get

$$\frac{d}{d\kappa}\int_\kappa T^{\mu\nu}\kappa_\nu\,d\Omega + \frac{d}{d\tau}\int_\kappa T^{\mu\nu}\frac{s_\nu}{\kappa}\,d\Omega = 0. \tag{19}$$

As $\kappa \to 0$ the second term in (19) becomes a perfect differential, and hence if there are any singular terms in the first term of (19) these must be perfect differentials and be compensated identically by those in the second term. (19) is the exact statement that the singular part of the inflow is a perfect differential. In case the energy tensor can be expanded as a series in ascending powers of κ for sufficiently small values of κ starting with some negative power of κ, then (19) shows that all the terms except the term independent of κ in the inflow must be perfect differentials.*

The singular terms being perfect differentials can always be compensated by the addition of suitable terms to A^μ in (9) and therefore *play no part in determining the motion of the point-particle*. The term independent of κ is not a perfect differential, and by putting it equal to a suitable perfect differential as in (9) the equation of translational motion of the point-particle is obtained. It is seen that to determine the equation of a point-particle it is only necessary to calculate that part of the inflow which remains finite as $\kappa \to 0$. The rest can be ignored.

It is clear that the same argument holds if two of the boundaries of the volume V are taken as any two surfaces passing through the points τ and $\tau + d\tau$ instead of the light cones. The argument also holds for any other tensor which satisfies a conservation law of the type (3), and therefore for the angular-momentum tensor of the field. In finding the rotational equations of the particle it is therefore only necessary to calculate the finite part of the inflow of angular momentum and to put this equal to a suitable perfect differential.

2. Let $X^{\mu\nu}$ be a tensor antisymmetric in μ and ν. Let us suppose that the coordinates ζ^0, ζ^1, ζ^2 and ζ^3 of the last section are so chosen that $\zeta^1 = \beta$ with constant β

---

* This result is implicit in Mathisson's paper (1940) and can be derived from his variational equations, but it is much more cumbersome and indirect to prove it by his method.

represents a closed two-dimensional surface ω forming the boundary of a portion S' of the three-dimensional surface $\zeta^0 = \alpha$, and assume that $\zeta^1 = \beta - \Delta\beta$ ($\Delta\beta$ positive) is a two-dimensional surface lying entirely within the three-dimensional surface so enclosed. Then it is well known that

$$\int_{S'} \frac{\partial X^{\nu\sigma}}{\partial x^\sigma}\, dS_\nu \equiv \int_{S'} \frac{\partial X^{\nu\sigma}}{\partial x^\sigma}\frac{\partial \zeta^0}{\partial x^\nu}\, |D|^{-1} d\zeta^1 d\zeta^2 d\zeta^3 = \int_\omega X^{\nu\sigma}\frac{\partial \zeta^0}{\partial x^\nu}\frac{\partial \zeta^1}{\partial x^\sigma}\, |D|^{-1} d\zeta^2 d\zeta^3. \qquad (20)$$

In this formula let us insert a tensor $K^{\mu\nu\sigma}$ antisymmetric in ν and σ in place of $X^{\nu\sigma}$, and take for the surface ζ^0 the surface of the tube $\kappa = \epsilon$ between the points τ_1 and τ_2. It is convenient in this case to take as the variables ζ, $\zeta^0 = \kappa$, $\zeta^1 = \tau$, $\zeta^2 = s^2$, $\zeta^3 = s^3$, and we obtain, as in the last section, remembering (11)–(15),

$$\begin{aligned}\int_{\tau_1}^{\tau_2}\int \frac{\partial K^{\mu\nu\sigma}}{\partial x^\sigma}\kappa_\nu\, d\Omega\, d\tau &= \int_\kappa K^{\mu\nu\sigma}\kappa_\nu \frac{s_\sigma}{\kappa}\, d\Omega\,\Big|_{\tau_1}^{\tau_2} \\ &= \int_\kappa K^{\mu\nu\sigma} v_\nu \frac{s_\sigma}{\kappa}\, d\Omega\,\Big|_{\tau_1}^{\tau_2}, \qquad (21)\end{aligned}$$

on account of the antisymmetry of $K^{\mu\nu\sigma}$ in ν and σ. Differentiating this with respect to τ,

$$\int_\kappa \frac{\partial K^{\mu\nu\sigma}}{\partial x^\sigma}\kappa_\nu\, d\Omega = \frac{d}{d\tau}\int_\kappa K^{\mu\nu\sigma} v_\nu \frac{s_\sigma}{\kappa}\, d\Omega. \qquad (22)$$

The inflow of $\partial K^{\mu\nu\sigma}/\partial x^\sigma$ over an infinitesimal portion of a world-tube of any arbitrary shape can similarly always be shown to be identically a perfect differential. Hence, if the energy-momentum tensor $T^{\mu\nu}$ is replaced by a new one $T'^{\mu\nu}$ defined by

$$T'^{\mu\nu} = T^{\mu\nu} + \frac{\partial K^{\mu\nu\sigma}}{\partial x^\sigma}, \qquad (23)$$

which obviously also satisfies a conservation equation like (3), then the inflow is only altered by the addition of perfect differentials, and these do not alter the equations of motion since they can be compensated by the addition of suitable terms to A^μ in (9).

Denote by $\dot{B}^\mu$ that part of the inflow, calculated by using the original tensor $T^{\mu\nu}$, which is a perfect differential. B^μ depends only on the variables at τ. Then taking

$$K^{\mu\nu\sigma} = \frac{1}{4\pi}\frac{B^\mu}{\kappa^3}(v^\nu s^\sigma - v^\sigma s^\nu) \qquad (24)$$

and inserting it into (22), we get, remembering (1), (2),

$$\int_\kappa \frac{\partial K^{\mu\nu\sigma}}{\partial x^\sigma}\kappa_\nu\, d\Omega = -\dot{B}^\mu. \qquad (25)$$

Since it has been shown that the singular parts of the inflow are always perfect differentials they can be included in B^μ, and it has therefore been proved that it is always possible to find a tensor $K^{\mu\nu\sigma}$ such that the inflow calculated from the modified tensor (23) is finite over the particular world-tube chosen.

The tensor (24) is by no means uniquely defined. Thus

$$K^{\mu\nu\sigma} = \frac{1}{4\pi}\frac{B^{\mu}}{\kappa^3}e^{-b\kappa^n}(v^{\nu}s^{\sigma}-v^{\sigma}s^{\nu}) \tag{26}$$

could equally well have been taken, with b a positive constant and n a number greater than the highest order of singularity in B^{μ}. This tensor would in fact have the advantage over (24) that its integral over any surface at infinity would always vanish.

The integral of the modified tensor over any arbitrary three-dimensional surface is not necessarily finite, nor is the inflow finite over a tube of any other shape than the one chosen above. The general properties of the energy tensor $T^{\mu\nu}$ are investigated in more detail in the paper by one of us (H.-C.) which follows this, and it is shown there that it is always possible to find a tensor $K^{\mu\nu\sigma}$ such that the modified tensor has no singularities of order higher than the third.

REFERENCES

Bhabha 1939 *Proc. Roy. Soc.* A, **172**, 384–409.
Bhabha 1940 *Proc. Ind. Acad. Sci.* A, **11**, 247–267, 467.
Bhabha 1941 *Proc. Roy. Soc.* A, **178**, 314–350.
Bhabha & Corben 1941 *Proc. Roy. Soc.* A, **178**, 273–314.
Dirac 1938 *Proc. Roy. Soc.* A, **167**, 148–169.
Mathisson 1940 *Proc. Camb. Phil. Soc.* **36**, 331–350.
Pryce 1938 *Proc. Roy. Soc.* A, **168**, 389–401.

On the removal of the infinite self-energies of point-particles

By Harish-Chandra, *J. H. Bhabha Memorial Student*
Cosmic Ray Research Unit, Indian Institute of Science, Bangalore

(*Communicated by H. J. Bhabha, F.R.S.—Received* 20 *March* 1944)

A general method is set up for modifying the energy-momentum tensor so as to remove the singularities in the flow of energy and momentum into the world-line of a particle without affecting the equations of motion of the particle. It is shown how the singularities of different order may be removed one by one.

In the case of the electromagnetic and meson fields it is shown that the modified tensor leads to a finite integral of energy and momentum over any space-like surface. In other cases the corresponding result may be secured by making a further modification in the tensor.

1. It has been shown in the preceding paper* that all singular terms in the expression for the rate of inflow† of energy and momentum are perfect differentials, and therefore the equations of motion of a point-particle calculated from considerations of the conservation of energy and momentum are always finite. One would now expect to be able to alter, without disturbing the equations of motion, the energy-momentum tensor $T^{\mu\nu}$ of the field so as to make the inflow finite, thus avoiding completely the appearance of infinities which are entirely spurious, at any rate, from the point of view of the equations of motion. It would be still better if it could be arranged that the total energy and momentum calculated from the modified tensor $\tilde{T}^{\mu\nu}$ were finite. This of course would automatically secure a finite inflow.

It is proved in A that the replacement of the energy-momentum tensor $T^{\mu\nu}$ by

$$\tilde{T}^{\mu\nu} = T^{\mu\nu} + \frac{\partial K^{\mu\nu\sigma}}{\partial x^\sigma}, \tag{1}$$

where $K^{\mu\nu\sigma}$ is antisymmetric in (ν, σ), does not alter the equations of motion. Therefore an attempt should be made to alter $T^{\mu\nu}$ in accordance with (1) in trying to get rid of the infinities. This method has already been used by Pryce (1938) to make the energy and momentum of the field finite in the case of the point electron. It will be shown quite generally in this paper that given an energy-momentum tensor $T^{\mu\nu}$ satisfying the conservation equation

$$\frac{\partial T^{\mu\nu}}{\partial x^\nu} = 0, \tag{2}$$

a tensor $K^{\mu\nu\sigma}$ can always be found so as to make the energy-momentum integrals calculated from $\tilde{T}^{\mu\nu}$ finite provided $T^{\mu\nu}$ fulfils certain very general restrictions. In general, $K^{\mu\nu\sigma}$ is not symmetric in μ, ν even when $T^{\mu\nu}$ is so, and therefore the modified tensor $\tilde{T}^{\mu\nu}$ is also not symmetric. But it will be shown that in the case of an electromagnetic charge or a dipole this symmetry can be achieved. However, the symmetry of the energy-momentum tensor, which is needed only for the purpose of constructing an angular-momentum tensor, is not necessary here, because it will be shown that a similar process can be applied to the angular-momentum tensor also, so that the total angular momentum calculated from this modified tensor is finite.

---

* Bhabha & Harish-Chandra (1944), referred to in this paper as A.

† This term is used in the same sense as in A.

2. Keeping to the notation of A it is now necessary to determine $K^{\mu\nu\sigma}$. Equations (18) and (19) of A are

$$\frac{d}{d\tau}\left\{\int_{\epsilon}^{\eta} d\kappa \int_{\kappa} T^{\mu\nu}\frac{s_\nu}{\kappa}\,d\Omega\right\} = \int_{\kappa=\epsilon} T^{\mu\nu}\kappa_\nu\,d\Omega - \int_{\kappa=\eta} T^{\mu\nu}\kappa_\nu\,d\Omega, \tag{3}$$

$$\frac{d}{d\tau}\int_{\kappa} T^{\mu\nu}\frac{s_\nu}{\kappa}\,d\Omega = -\frac{d}{d\kappa}\int_{\kappa} T^{\mu\nu}\kappa_\nu\,d\Omega. \tag{4}$$

Here $\int_\kappa$ denotes integration over the sphere of intersection of the light cone at τ with the tube $\kappa =$ constant. This sphere will be called the 'retarded sphere' of radius κ at τ. In the appendix (equation (98)) it is shown that

$$\kappa\frac{d}{d\kappa}\int_{\kappa} T^{\mu\nu}\kappa_\nu\frac{s^\alpha}{\kappa}\frac{s^\beta}{\kappa}\cdots\frac{s^\lambda}{\kappa}\,d\Omega = \int_{\kappa}\left\{\frac{\partial}{\partial x^\sigma}(T^{\mu\nu}s^\sigma - T^{\mu\sigma}s^\nu)\right\}\kappa_\nu\frac{s^\alpha}{\kappa}\frac{s^\beta}{\kappa}\cdots\frac{s^\lambda}{\kappa}\,d\Omega, \tag{5}$$

where $\alpha, \beta, \ldots, \lambda$ are any number of free indices. Now introduce two operators D and θ defined in the following way:

$$D = \kappa\frac{d}{d\kappa}, \tag{6a}$$

$$\begin{aligned}\theta U^{\alpha\beta\ldots\lambda\nu} &= \frac{\partial}{\partial x^\sigma}(U^{\alpha\beta\ldots\lambda\nu}s^\sigma - U^{\alpha\beta\ldots\lambda\sigma}s^\nu)\\ &= 2U^{\alpha\beta\ldots\nu} + U^{\alpha\beta\ldots\sigma}\frac{s_\sigma}{\kappa}v^\nu + s^\sigma\frac{\partial U^{\alpha\beta\ldots\nu}}{\partial x^\sigma} - s^\nu\frac{\partial U^{\alpha\beta\ldots\sigma}}{\partial x^\sigma},\end{aligned} \tag{6b}$$

where $U^{\alpha\beta\ldots\lambda\nu}$ is any arbitrary tensor with any number of indices. It should be noted that

$$\frac{\partial}{\partial x^\nu}(\theta U^{\alpha\beta\ldots\lambda\nu}) = 0 \tag{7}$$

owing to the antisymmetry in ν, σ of the expression inside the brackets in (6b). Obviously both D and θ are linear operators. Also $D\kappa^n = n\kappa^n$ and therefore by repetition $D^r\kappa^n = n^r\kappa^n$, so that if $f(D)$ is a polynomial expressible in powers of D, $f(D)\,\kappa^n = f(n)\,\kappa^n$. This holds whether n is positive or negative.

Using D and θ, (5) can be written in the form

$$D\int_{\kappa} T^{\mu\nu}\kappa_\nu\frac{s^\alpha}{\kappa}\cdots\frac{s^\lambda}{\kappa}\,d\Omega = \int_{\kappa}\theta T^{\mu\nu}.\,\kappa_\nu\frac{s^\alpha}{\kappa}\cdots\frac{s^\lambda}{\kappa}\,d\Omega.$$

Since $\theta T^{\mu\nu}$ is conserved with respect to ν as $T^{\mu\nu}$ was, the same considerations as above can be applied to $\theta T^{\mu\nu}$ instead of $T^{\mu\nu}$ and hence

$$D^2\int_\kappa T^{\mu\nu}\kappa_\nu\frac{s^\alpha}{\kappa}\dots\frac{s^\lambda}{\kappa}d\Omega = D\int_\kappa (\theta T^{\mu\nu})\kappa_\nu\frac{s^\alpha}{\kappa}\dots\frac{s^\lambda}{\kappa}d\Omega = \int_\kappa \theta^2 T^{\mu\nu}.\kappa_\nu\frac{s^\alpha}{\kappa}\dots\frac{s^\lambda}{\kappa}d\Omega.$$

By repetition it can be proved that

$$D^r\int_\kappa T^{\mu\nu}\kappa_\nu\frac{s^\alpha}{\kappa}\dots\frac{s^\lambda}{\kappa}d\Omega = \int_\kappa \theta^r T^{\mu\nu}.\kappa_\nu\frac{s^\alpha}{\kappa}\dots\frac{s^\lambda}{\kappa}d\Omega,$$

and therefore
$$f(D)\int_\kappa T^{\mu\nu}\kappa_\nu\frac{s^\alpha}{\kappa}\dots\frac{s^\lambda}{\kappa}d\Omega = \int_\kappa f(\theta)\,T^{\mu\nu}.\kappa_\nu\frac{s^\alpha}{\kappa}\dots\frac{s^\lambda}{\kappa}d\Omega, \tag{8}$$

where $f(D)$ is a polynomial expressible in positive integral powers of D.

It will now be assumed that the part of $T^{\mu\nu}$ which becomes infinite as $\kappa\to 0$ can be expressed for sufficiently small values of κ in the form of a series involving a finite number of negative powers of κ, so that for sufficiently small κ one can write*

$$T^{\mu\nu} = T_0^{\mu\nu} + \sum_{r=1}^{r=m+2}\frac{a_{(-r)}^{\mu\nu}}{\kappa^r}, \tag{9}$$

where $T_0^{\mu\nu}$ remains finite as $\kappa\to 0$ and $a_{(-r)}^{\mu\nu}$ are functions of s^α/κ and τ only. It is assumed that these functions are finite and continuous. That the energy-momentum tensors of the electromagnetic and the meson fields actually satisfy these requirements is shown in the appendix. Therefore

$$\int_\kappa T^{\mu\nu}\kappa_\nu\frac{s^\alpha}{\kappa}\dots\frac{s^\lambda}{\kappa}d\Omega = c_0^{\mu\alpha\dots\lambda} + \sum_{r=1}^{r=m}\frac{c_{(-r)}^{\mu\alpha\dots\lambda}}{\kappa^r} \tag{10}$$

can be written, where $c_0^{\mu\alpha\dots\lambda}$ is a function of κ and τ which remains finite as $\kappa\to 0$ and $c_{(-r)}^{\mu\alpha\dots\lambda}$ are functions of τ only. It is assumed that for sufficiently small κ, $c_0^{\mu\alpha\dots\lambda}$ is expressible as a series in positive powers of κ. This also is true for the electromagnetic and meson fields.

Applying the operator $f(D)$ to both the sides of (10),

$$f(D)\int_\kappa T^{\mu\nu}\kappa_\nu\frac{s^\alpha}{\kappa}\dots\frac{s^\lambda}{\kappa}d\Omega = f(D)\,c_0^{\mu\alpha\dots\lambda} + \sum_{r=1}^{r=m}f(-r)\frac{c_{(-r)}^{\mu\alpha\dots\lambda}}{\kappa^r}. \tag{11}$$

Taking
$$f(D) = \prod_{k=1}^{k=m}\left(1+\frac{D}{k}\right), \tag{12}$$
we get $f(-r) = 0$ for $1\leqslant r\leqslant m$.

From (11) and (8) it follows that

$$\prod_{k=1}^{k=m}\left(1+\frac{D}{k}\right)\int_\kappa T^{\mu\nu}\kappa_\nu\frac{s^\alpha}{k}\dots\frac{s^\lambda}{\kappa}d\Omega = \int_\kappa\left\{\prod_{k=1}^{k=m}\left(1+\frac{\theta}{k}\right)T^{\mu\nu}\right\}\kappa_\nu\frac{s^\alpha}{\kappa}\dots\frac{s^\lambda}{\kappa}d\Omega$$
$$= \prod_{k=1}^{k=m}\left(1+\frac{D}{k}\right)c_0^{\mu\alpha\dots\lambda}. \tag{13}$$

---

* The highest singularity is written as of the order $m+2$ for convenience. The removal of only those singularities from $T^{\mu\nu}$ is necessary which are of an order higher than the 2nd, since singularities of the 2nd or lesser order give a finite energy integral (Pryce 1938).

Since the right side of this equation is non-singular, the left must be so, i.e. $\prod_{k=1}^{k=m}\left(1+\frac{\theta}{k}\right)T^{\mu\nu}$ gives a finite integral.

From this it follows that
$$\left\{\prod_{k=1}^{k=m}\left(1+\frac{\theta}{k}\right)T^{\mu\nu}\right\}\kappa_\nu \tag{14}$$
cannot contain any singularity of an order higher than the 2nd, because if it contained a term of the type b^μ/κ^{r+2} (where $r>0$), this would give after integration a singular term of the type $c^{\mu\alpha\ldots\lambda}/\kappa^r$, where
$$c^{\mu\alpha\ldots\lambda}=\int b^\mu\frac{s^\alpha}{\kappa}\ldots\frac{s^\lambda}{\kappa}d\omega \quad \left[d\omega=\frac{d\Omega}{\kappa^2}=\text{the element of solid angle}\right].$$
Since such a term is absent from the right side of (13) it follows that
$$c^{\mu\alpha\ldots\lambda}=\int b^\mu\frac{s^\alpha}{\kappa}\ldots\frac{s^\lambda}{\kappa}d\omega=0.$$
As in the rest system s^α/κ are nothing else than the direction cosines (except s^0/κ which is 1), it follows from the expansibility of every continuous function defined on the surface of a sphere, in terms of surface harmonics, that $b^\mu=0$. Thus (14) does not contain any singularities higher than of the 2nd order.

Because
$$\prod_{k=1}^{k=m}\left(1+\frac{\theta}{k}\right)=1+a\theta+b\theta^2+\ldots=1+\theta(a+b\theta+\ldots), \tag{15}$$
where a, b, ... are numerical coefficients, it is clear that
$$\prod_{k=1}^{k=m}\left(1+\frac{\theta}{k}\right)T^{\mu\nu}=T^{\mu\nu}+\theta(a+b\theta+\ldots)\,T^{\mu\nu}=T^{\mu\nu}+\theta K^{\mu\nu},$$
where
$$K^{\mu\nu}\equiv(a+b\theta+\ldots)\,T^{\mu\nu}.$$
Further, as $\theta K^{\mu\nu}=\partial K^{\mu\nu\sigma}/\partial x^\sigma$, where $K^{\mu\nu\sigma}=K^{\mu\nu}s^\sigma-K^{\mu\sigma}s^\nu$, the modified tensor
$$\prod_{k=1}^{k=m}\left(1+\frac{\theta}{k}\right)T^{\mu\nu}=T^{\mu\nu}+\frac{\partial K^{\mu\nu\sigma}}{\partial x^\sigma},$$
which is just of the required form. Put
$$\prod_{k=1}^{k=m}\left(1+\frac{\theta}{k}\right)T^{\mu\nu}=T'^{\mu\nu}. \tag{16}$$

It should be noted in passing that it can be shown directly that $T'^{\mu\nu}$ does not contain singularities of an order higher than the 3rd, because, as shown in the appendix (equation (102)),
$$D\int_\kappa T^{\mu\sigma}(\delta^\nu_\sigma-v_\sigma v^\nu)\frac{s^\alpha}{\kappa}\ldots\frac{s^\lambda}{\kappa}d\Omega=\int_\kappa \theta T^{\mu\sigma}.(\delta^\nu_\sigma-v_\sigma v^\nu)\frac{s^\alpha}{\kappa}\ldots\frac{s^\lambda}{\kappa}d\Omega.$$
Hence
$$\int_\kappa T'^{\mu\sigma}(\delta^\nu_\sigma-v_\sigma v^\nu)\frac{s^\alpha}{\kappa}\ldots\frac{s^\lambda}{\kappa}d\Omega=\prod_{k=1}^{k=m}\left(1+\frac{D}{k}\right)\int_\kappa T^{\mu\sigma}(\delta^\nu_\sigma-v_\sigma v^\nu)\frac{s^\alpha}{\kappa}\ldots\frac{s^\lambda}{\kappa}d\Omega. \tag{17}$$

Through the same reasoning as used before it can be shown that, since the right side of (17) is non-singular, so also must be the left. From this it can be inferred exactly as in the case of $T'^{\mu\nu}\kappa_\nu$ that $T'^{\mu\sigma}(\delta^\nu_\sigma - v_\sigma v^\nu)$ does not contain singularities of an order higher than the 2nd.

It is further found (appendix equation (100)) that

$$D\int_\kappa T^{\mu\nu}s_\nu \frac{s^\alpha}{\kappa}\dots\frac{s^\lambda}{\kappa}d\Omega = \int_\kappa \theta T^{\mu\nu}.s_\nu\frac{s^\alpha}{\kappa}\dots\frac{s^\lambda}{\kappa}d\Omega.$$

Hence
$$\int_\kappa T'^{\mu\nu}s_\nu\frac{s^\alpha}{\kappa}\dots\frac{s^\lambda}{\kappa}d\Omega = \prod_{k=1}^{k=m}\left(1+\frac{D}{k}\right)\int_\kappa T^{\mu\nu}s_\nu\frac{s^\alpha}{\kappa}\dots\frac{s^\lambda}{\kappa}d\Omega. \tag{18}$$

From this one infers precisely as before that $T'^{\mu\nu}s_\nu$ does not contain singularities higher than of the 2nd order, so that $T'^{\mu\nu}\frac{s_\nu}{\kappa}$ or $T'^{\mu\nu}\frac{s_\nu}{\kappa}(1-\kappa')$ cannot contain singularities higher than $1/\kappa^3$. Since it has also been proved that

$$T'^{\mu\nu}\kappa_\nu = T'^{\mu\nu}\left\{v_\nu - \frac{s_\nu}{\kappa}(1-\kappa')\right\}$$

does not have singularities higher than of the 2nd order, it follows that $T'^{\mu\nu}v_\nu$ does not possess terms of order higher than $1/\kappa^3$. Now write $T'^{\mu\nu}$ in the form

$$T'^{\mu\nu} = T'^{\mu\sigma}(\delta^\nu_\sigma - v_\sigma v^\nu) + (T'^{\mu\sigma}v_\sigma)\,v^\nu.$$

The first term does not contain terms of order higher than $1/\kappa^2$ and the second of order higher than $1/\kappa^3$. So the highest singularity in $T'^{\mu\nu}$ can at most be of the order $1/\kappa^3$.

3. We now proceed to examine in detail the process of removal of the singular terms of $T^{\mu\nu}$ by the operator $\prod_{k=1}^{k=m}\left(1+\frac{\theta}{k}\right)$. In addition to θ we define another operator ϕ by the equation

$$\phi U^{\alpha\dots\nu} = 2U^{\alpha\dots\nu} + s^\sigma\frac{\partial U^{\alpha\dots\sigma}}{\partial x^\sigma} + U^{\alpha\dots\sigma}\frac{s_\sigma}{\kappa}v^\nu, \tag{19}$$

so that from (6b) and (19) is obtained

$$\theta U^{\alpha\dots\nu} = \phi U^{\alpha\dots\nu} - s^\nu\frac{\partial U^{\alpha\dots\sigma}}{\partial x^\sigma}. \tag{20}$$

When $U^{\alpha\dots\sigma}$ is conserved with respect to σ,

$$\theta U^{\alpha\dots\nu} = \phi U^{\alpha\dots\nu}. \tag{21}$$

While $\theta U^{\alpha\dots\nu}$ is always conserved with respect to ν, $\phi U^{\alpha\dots\nu}$ is conserved in general only when $U^{\alpha\dots\nu}$ itself is conserved.

Let $T^{\mu\nu}_{(-n)}$ denote the nth order term in $T^{\mu\nu}$ so that in accordance with (9) $T^{\mu\nu}_{(-n)} = \frac{a^{\mu\nu}_{(-n)}}{\kappa^n}$. Then

$$\phi T^{\mu\nu}_{-n)} = 2T^{\mu\nu}_{(-n)} + s^\sigma\frac{\partial T^{\mu\nu}_{(-n)}}{\partial x^\sigma} + T^{\mu\nu}_{(-n)}\frac{s_\sigma}{\kappa}v^\nu.$$

As before assume $a^{\mu\nu}_{(-n)}$ to be a function of the arguments s^α/κ and τ only. Now both τ and s^α/κ are to be treated as constants with respect to the operator $s^\sigma\dfrac{\partial}{\partial x^\sigma}$, because $s^\sigma\dfrac{\partial\tau}{\partial x^\sigma}=0$ and $s^\sigma\dfrac{\partial}{\partial x^\sigma}\left(\dfrac{s^\alpha}{\kappa}\right)=0$, so that $s^\sigma\dfrac{\partial a^{\mu\nu}_{(-n)}}{\partial x^\sigma}=0$. Therefore

$$s^\sigma\frac{\partial T^{\mu\nu}_{(-n)}}{\partial x^\sigma}=-n\frac{a^{\mu\nu}_{(-n)}}{\kappa^{n+1}}\kappa_\sigma s^\sigma=-n\frac{a^{\mu\nu}_{(-n)}}{\kappa^n}=-nT^{\mu\nu}_{(-n)}. \tag{22}$$

Thus
$$\phi T^{\mu\nu}_{(-n)}=(2-n)\,T^{\mu\nu}_{(-n)}+T^{\mu\sigma}_{(-n)}\frac{s_\sigma}{\kappa}v^\nu=\left[(2-n)\,\delta^\nu_\sigma+v^\nu\frac{s_\sigma}{\kappa}\right]T^{\mu\sigma}_{(-n)}. \tag{23}$$

Now since $s^\rho\dfrac{\partial}{\partial x^\rho}\left(\dfrac{s_\sigma}{\kappa}v^\nu\right)=0$, it follows that

$$\phi^2T^{\mu\nu}_{(-n)}=\left[(2-n)\,\delta^\nu_\rho+v^\nu\frac{s_\rho}{\kappa}\right]\left[(2-n)\,\delta^\rho_\sigma+v^\rho\frac{s_\sigma}{\kappa}\right]T^{\mu\sigma}_{(-n)}.$$

By repetition
$$\phi^lT^{\mu\nu}_{(-n)}=\left(\left[(2-n)\,\delta+v\frac{s}{\kappa}\right]^l\right)^\nu_\sigma T^{\mu\sigma}_{(-n)} \tag{24}$$

for any positive integer l, where

$$\left(\left[(2-n)\,\delta+v\frac{s}{\kappa}\right]^l\right)^\nu_\sigma$$
$$\equiv\left[(2-n)\,\delta^\nu_{\rho_1}+v^\nu\frac{s_{\rho_1}}{\kappa}\right]\left[(2-n)\,\delta^{\rho_1}_{\rho_2}+v^{\rho_1}\frac{s_{\rho_2}}{\kappa}\right]\dots\left[(2-n)\,\delta^{\rho_{l-1}}_\sigma+v^{\rho_{l-1}}\frac{s_\sigma}{\kappa}\right].$$

Put
$$a\delta^\nu_\sigma+bv^\nu\frac{s_\sigma}{\kappa}\equiv\Big[A(a,b)\Big]^\nu_\sigma \tag{25a}$$

and
$$\left[\left(a\delta+bv\frac{s}{\kappa}\right)^l\right]^\nu_\sigma\equiv\Big[A^l(a,b)\Big]^\nu_\sigma. \tag{25b}$$

The indices will sometimes be omitted. For example, we shall write

$$\phi^lT_{(-n)}=A^l(2-n,1)\,T_{(-n)}, \tag{26a}$$

meaning thereby
$$\phi^lT^{\mu\nu}_{(-n)}=\Big[A^l(2-n,1)\Big]^\nu_\sigma T^{\mu\sigma}_{(-n)}. \tag{26b}$$

It is also seen that
$$cA(a,b)=A(ca,cb), \tag{27a}$$

$$A(a_1,b_1)+A(a_2,b_2)=A(a_1+a_2,b_1+b_2), \tag{27b}$$

and
$$c=A(c,0), \tag{27c}$$

where c is an ordinary number. On using (26) and (27)

$$\prod_{k=1}^{k=m}\left(1+\frac{\phi}{k}\right)T_{(-n)}=\prod_{k=1}^{k=m}\left[1+\frac{A(2-n,1)}{k}\right]T_{(-n)}=\prod_{k=1}^{k=m}A\left(1+\frac{2-n}{k},\frac{1}{k}\right)T_{(-n)}. \tag{28}$$

Consider now $A(a_1, b_1)\, A(a_2, b_2)$. It is found that

$$\Big[A(a_1, b_1)\, A(a_2, b_2)\Big]^{\nu}_{\sigma} = \left(a_1 \delta^{\nu}_{\rho} + b_1 v^{\nu} \frac{s_\rho}{\kappa}\right)\left(a_2 \delta^{\rho}_{\sigma} + b_2 v^{\rho} \frac{s_\sigma}{\kappa}\right)$$

$$= a_1 a_2 \delta^{\nu}_{\sigma} + [(a_1 + b_1)(a_2 + b_2) - a_1 a_2]\, v^{\nu} \frac{s_\sigma}{\kappa}$$

$$= \Big[A(a_1 a_2, [a_1 + b_1][a_2 + b_2] - a_1 a_2)\Big]^{\nu}_{\sigma}.$$

It can easily be proved by induction that

$$A(a_1, b_1)\, A(a_2, b_2) \dots A(a_n, b_n)$$
$$= A(a_1 a_2 \dots a_n, [a_1 + b_1][a_2 + b_2] \dots [a_n + b_n] - a_1 a_2 \dots a_n). \tag{29}$$

Thus we get immediately

$$\prod_{k=1}^{k=m} A\left(1 + \frac{2-n}{k}, \frac{1}{k}\right) = A\left(\prod_{k=1}^{k=m}\left(1 + \frac{2-n}{k}\right), \prod_{k=1}^{k=m}\left(1 + \frac{3-n}{k}\right) - \prod_{k=1}^{k=m}\left(1 + \frac{2-n}{k}\right)\right). \tag{30}$$

Hence if $m \geqslant n-2$ and $n \geqslant 4$,

$$\prod_{k=1}^{k=m} A\left(1 + \frac{2-n}{k}, \frac{1}{k}\right) = A(0, 0) = 0.$$

Thus
$$\prod_{k=1}^{k=m}\left(1 + \frac{\phi}{k}\right) T_{(-n)} = 0 \quad \text{for} \quad m+2 \geqslant n \geqslant 4.$$

Since $m+2$ has already been chosen equal to the order of the highest singularity in $T^{\mu\nu}$, it follows that the operator $\prod_{k=1}^{k=m}\left(1 + \frac{\phi}{k}\right)$ when applied to $T^{\mu\nu}$ removes all singular terms higher than the 3rd. For the term of the 3rd order, from (28) and (30),

$$\prod_{k=1}^{k=m}\left(1 + \frac{\phi}{k}\right) T_{(-3)} = A(0, 1)\, T_{(-3)} = v^{\nu} \frac{s_\sigma}{\kappa} T^{\mu\sigma}_{(-3)}.$$

Thus
$$\prod_{k=1}^{k=m}\left(1 + \frac{\phi}{k}\right) T^{\mu\nu} = \prod_{k=1}^{k=m}\left(1 + \frac{\phi}{k}\right) T^{\mu\nu}_{-2)} + T^{\mu\sigma}_{(-3)} \frac{s_\sigma}{\kappa} v^{\nu},$$

where $T^{\mu\nu}_{-2)}$ denotes the part of $T^{\mu\nu}$ remaining after omitting terms of 3rd order or higher.*

---

* $T^{\mu\nu}_{-n)}$ denotes the part of $T^{\mu\nu}$ left over after omitting terms of order higher than the nth, and $T^{\mu\nu}_{(-n}$ the part obtained by retaining only terms of the nth or higher order, so that

$$T^{\mu\nu}_{-n+1)} + T^{\mu\nu}_{(-n} = T^{\mu\nu}.$$

The suffixes $-n)$ and $(-n$ attached to any other quantity have analogous significance. Similarly the suffix $(-n)$ always denotes the term of the nth order.

Since $T^{\mu\nu}$ is conserved with respect to ν, $\phi T^{\mu\nu} = \theta T^{\mu\nu}$ and $\phi^l T^{\mu\nu} = \theta^l T^{\mu\nu}$, so that

$$T'^{\mu\nu} = \prod_{k=1}^{k=m}\left(1+\frac{\theta}{k}\right) T^{\mu\nu} = \prod_{k=1}^{k=m}\left(1+\frac{\phi}{k}\right) T^{\mu\nu} = \prod_{k=1}^{k=m}\left(1+\frac{\phi}{k}\right) T^{\mu\nu}_{-2)} + T^{\mu\sigma}_{(-3)}\frac{s_\sigma}{\kappa}v^\nu, \tag{31}$$

which is in conformity with the proved fact that $T'^{\mu\nu}$ does not contain singularities higher than of the 3rd order.

In accordance with (15)

$$\prod_{k=1}^{k=m}\left(1+\frac{\phi}{k}\right) = 1+\phi(a+b\phi+\ldots),$$

so that

$$\begin{aligned} T'^{\mu\nu} &= T^{\mu\nu}_{-2)} + \phi(a+b\phi+\ldots)\, T^{\mu\nu}_{-2)} + T^{\mu\sigma}_{(-3)}\frac{s_\sigma}{\kappa}v^\nu \\ &= T^{\mu\nu}_{-2)} + \theta(a+b\phi+\ldots)\, T^{\mu\nu}_{-2)} + T^{\mu\sigma}_{(-3)}\frac{s_\sigma}{\kappa}v^\nu + s^\nu\frac{\partial}{\partial x^\sigma}[(a+b\phi+\ldots)\, T^{\mu\sigma}_{-2)}]. \end{aligned} \tag{32}$$

A new tensor $T''^{\mu\nu}$ is now introduced, defined by

$$T''^{\mu\nu} \equiv T'^{\mu\nu} - \theta(a+b\phi+\ldots)\, T^{\mu\nu}_{-2)} = T^{\mu\nu}_{-2)} + T^{\mu\sigma}_{(-3)}\frac{s_\sigma}{\kappa}v^\nu + s^\nu\frac{\partial}{\partial x^\sigma}[(a+b\phi+\ldots)\, T^{\mu\sigma}_{-2)}]. \tag{33}$$

$T''^{\mu\nu}$ is obviously conserved and is of the form $T^{\mu\nu}+\dfrac{\partial K^{\mu\nu\sigma}}{\partial x^\sigma}$, where $K^{\mu\nu\sigma}$ is antisymmetrical in ν,σ and therefore $T''^{\mu\nu}$ instead of $T'^{\mu\nu}$ may be taken as the modified tensor.

It will now be shown that

$$\frac{\partial}{\partial x^\sigma}[(a+b\phi+\ldots)\, T^{\mu\sigma}_{-2)}] = a\left[\frac{\partial T^{\mu\sigma}_{(-2)}}{\partial x^\sigma}\right]_{(-3)} = a\frac{\partial T^{\mu\sigma}_{-2)}}{\partial x^\sigma}. \tag{34}$$

It has been seen above that the order of a term is not changed by application of the operator ϕ. Hence $(a+b\phi+\ldots)\, T^{\mu\nu}_{-2)}$ contains no singularities higher than of the 2nd order. Since

$$T^{\mu\nu}_{(-3} = T^{\mu\nu} - T^{\mu\nu}_{-2)},$$

it follows from the conservation of $T^{\mu\sigma}$ and therefore of $(a+b\phi+\ldots)\, T^{\mu\sigma}$ that

$$\frac{\partial}{\partial x^\sigma}[(a+b\phi+\ldots)\, T^{\mu\sigma}_{-2)}] = -\frac{\partial}{\partial x^\sigma}[(a+b\phi+\ldots)\, T^{\mu\sigma}_{(-3}]. \tag{35}$$

Now

$$\begin{aligned} \frac{\partial T^{\mu\nu}_{(-n)}}{\partial x^\sigma} &= \frac{1}{\kappa^n}\left[\dot a^{\mu\nu}_{(-n)}\frac{s_\sigma}{\kappa} + a^{\mu\nu,\rho}_{(-n)}\frac{\partial}{\partial x^\sigma}\left(\frac{s_\rho}{\kappa}\right) - n\frac{a^{\mu\nu}_{(-n)}}{\kappa}\kappa_\sigma\right] \\ &= \frac{\dot a^{\mu\nu}_{(-n)}}{\kappa^n}\frac{s_\sigma}{\kappa} + \frac{a^{\mu\nu,\rho}_{(-n)}}{\kappa^{n+1}}\left\{g_{\rho\sigma} - \frac{s_\sigma}{\kappa}v_\rho - \frac{s_\rho}{\kappa}v_\sigma + \frac{s_\rho}{\kappa}\frac{s_\sigma}{\kappa}(1-\kappa')\right\} - n\frac{a^{\mu\nu}_{(-n)}}{\kappa^{n+1}}\left[v_\sigma - (1-\kappa')\frac{s_\sigma}{\kappa}\right], \end{aligned} \tag{36}$$

where

$$\dot a^{\mu\nu} = \frac{\partial a^{\mu\nu}_{(-n)}}{\partial\tau} \quad\text{and}\quad a^{\mu\nu,\rho} = \frac{\partial a^{\mu\nu}_{(-n)}}{\partial\left(\dfrac{s_\rho}{\kappa}\right)}, \tag{37}$$

treating $\tau, s_\rho/\kappa$ as independent variables. It follows therefore that on differentiation $T^{\mu\nu}_{(-n)}$ can give terms only of the nth and $n+1$th order. The same result is obviously applicable to terms of different orders in $(a+b\phi+\dots)\,T^{\mu\nu}$. Thus the highest singularity in $\frac{\partial}{\partial x^\sigma}[(a+b\phi+\dots)\,T^{\mu\nu}]$ is of the 3rd order, which is also the order of the lowest singularity in $\frac{\partial}{\partial x^\sigma}[(a+b\phi+\dots)\,T^{\mu\sigma}_{(-3)}]$. From equation (35) it follows therefore that $\frac{\partial}{\partial x^\sigma}[(a+b\phi+\dots)\,T^{\mu\sigma}_{(-2)}]$ contains only one term of the 3rd order, because the terms of lower order which it could contain do not exist on the right side of (35). Since only the 2nd order term in $(a+b\phi+\dots)\,T^{\mu\sigma}_{(-2)}$ can contribute to a 3rd order term in the divergence, it is sufficient to calculate the 2nd order term of $(a+b\phi+\dots)\,T^{\mu\sigma}_{(-2)}$ and find out the 3rd order term that results from it on differentiation.

Now proceed to determine $(a+b\phi+\dots)\,T^{\mu\nu}_{(-n)}$. From (26), (27) and (29) it follows immediately that

$$(a+b\phi+\dots)\,T_{(-n)} = A(p,q)\,T_{(-n)}, \tag{38}$$

where p, q are some numerical numbers. However, on using (27)–(30) one gets

$$\begin{aligned}\prod_{k=1}^{k=m}\left(1+\frac{\phi}{k}\right)T_{(-n)} &= [1+\phi(a+b\phi+\dots)]\,T_{(-n)} = T_{(-n)}+\phi A(p,q)\,T_{(-n)}\\ &= T_{(-n)}+A(2-n,1)\,A(p,q)\,T_{(-n)}\\ &= A(1+p(2-n),\ q(3-n)+p)\,T_{(-n)}\\ &= A\left[\prod_{k=1}^{k=m}\left(1+\frac{2-n}{k}\right),\ \prod_{k=1}^{k=m}\left(1+\frac{3-n}{k}\right)-\prod_{k=1}^{k=m}\left(1+\frac{2-n}{k}\right)\right]T_{(-n)}.\end{aligned}$$

Thus

$$1+p(2-n) = \prod_{k=1}^{k=m}\left(1+\frac{2-n}{k}\right),\quad q(3-n)+p = \prod_{k=1}^{k=m}\left(1+\frac{3-n}{k}\right)-\prod_{k=1}^{k=m}\left(1+\frac{2-n}{k}\right),$$

i.e.

$$p = \frac{1}{2-n}\left[\prod_{k=1}^{k=m}\left(1+\frac{2-n}{k}\right)-1\right], \tag{39a}$$

$$q = \frac{1}{3-n}\left[\prod_{k=1}^{k=m}\left(1+\frac{3-n}{k}\right)-\left(1+\frac{1}{2-n}\right)\prod_{k=1}^{k=m}\left(1+\frac{2-n}{k}\right)+\frac{1}{2-n}\right]. \tag{39b}$$

By passing to the limit $n \to 2$ it is easy to show that in this case $p = a = 1+\frac{1}{2}+\dots+\frac{1}{m}$ and $q = m-a$. Therefore

$$\frac{\partial}{\partial x^\sigma}[(a+b\phi+\dots)\,T^{\mu\sigma}_{(-2)}] = \left\{\frac{\partial}{\partial x^\sigma}\left[aT^{\mu\sigma}_{(-2)}+(m-a)\,T^{\mu\rho}_{(-2)}\frac{s_\rho}{\kappa}v^\sigma\right]\right\}_{(-3)}, \tag{40}$$

the suffix denoting that only the 3rd order term resulting after differentiation is to be retained. Now, since

$$\frac{\partial}{\partial x^\sigma}\left(\frac{s_\rho}{\kappa}v^\sigma\right) = 0, \tag{41}$$

it is found that

$$\frac{\partial}{\partial x^\sigma}\left[T^{\mu\rho}_{(-n)}\frac{s_\rho}{\kappa}v^\sigma\right]=\frac{s_\rho}{\kappa}v^\sigma\frac{\partial T^{\mu\nu}_{(-n)}}{\partial x^\sigma}=\frac{s_\rho}{\kappa}\left[\frac{\dot{a}^{\mu\rho}_{(-n)}}{\kappa^n}-\frac{a^{\mu\rho,\lambda}_{(-n)}s_\lambda}{\kappa^n}\frac{\kappa'}{\kappa}-n\frac{a^{\mu\rho}_{(-n)}}{\kappa^n}\frac{\kappa'}{\kappa}\right], \tag{42}$$

where the various symbols have the same meaning as in (37). Since $\dot{a}^{\mu\rho}_{(-n)}$, $a^{\mu\rho,\lambda}_{(-n)}$, $a^{\mu\rho}_{(-n)}$ are finite functions of s^α/κ and τ only, it is obvious that $\frac{\partial}{\partial x^\sigma}\left[T^{\mu\rho}_{(-n)}\frac{s_\rho}{\kappa}v^\sigma\right]$ is of the nth order. So that

$$\left\{\frac{\partial}{\partial x^\sigma}\left[T^{\mu\rho}_{(-2)}\frac{s_\rho}{\kappa}v^\sigma\right]\right\}_{(-3)}=0. \tag{43}$$

By comparing (36) and (42) it is also seen that

$$\frac{s_\rho}{\kappa}v^\sigma\frac{\partial T^{\mu\nu}_{(-n)}}{\partial x^\sigma}=\left[\frac{\partial T^{\mu\nu}_{(-n)}}{\partial x^\rho}\right]_{(-n)}, \tag{44}$$

a result which will be used later. So from (33), (40) and (43)

$$T''^{\mu\nu}=T^{\mu\nu}_{(-2)}+T^{\mu\sigma}_{(-3)}\frac{s_\sigma}{\kappa}v^\nu+s^\nu\left\{\frac{\partial T^{\mu\nu}_{(-2)}}{\partial x^\sigma}\right\}_{(-3)}a. \tag{45}$$

It should be observed that

$$\frac{\partial}{\partial x^\nu}\left[s^\nu\left\{\frac{\partial T^{\mu\sigma}_{(-2)}}{\partial x^\sigma}\right\}_{(-3)}\right]=3\left\{\frac{\partial T^{\mu\sigma}_{(-2)}}{\partial x^\sigma}\right\}_{(-3)}+s^\nu\frac{\partial}{\partial x^\nu}\left\{\frac{\partial T^{\mu\sigma}_{(-2)}}{\partial x^\sigma}\right\}_{(-3)}=0, \tag{46}$$

since $\left\{\frac{\partial T^{\mu\sigma}_{(-2)}}{\partial x^\sigma}\right\}_{(-3)}$ is of the 3rd order, and therefore by a relation analogous to (22)

$$s^\nu\frac{\partial}{\partial x^\nu}\left\{\frac{\partial T^{\mu\sigma}_{(-2)}}{\partial x^\sigma}\right\}_{(-3)}=-3\left\{\frac{\partial T^{\mu\sigma}_{(-2)}}{\partial x^\sigma}\right\}_{(-3)}.$$

Now since $T''^{\mu\nu}$ is conserved with respect to ν, as also $s^\nu\left\{\frac{\partial T^{\mu\sigma}_{(-2)}}{\partial x^\sigma}\right\}_{(-3)}$, it follows from (45) that

$$\frac{\partial T^{\mu\nu}_{(-2)}}{\partial x^\nu}+\frac{\partial}{\partial x^\nu}\left[T^{\mu\sigma}_{(-3)}\frac{s_\sigma}{\kappa}v^\nu\right]=0. \tag{47}$$

Further, using again the relation expressed by (22)

$$\begin{aligned}\theta\left[T^{\mu\sigma}_{(-3)}\frac{s_\sigma}{\kappa}v^\nu\right]&=-s^\nu\frac{\partial}{\partial x^\rho}\left[T^{\mu\sigma}_{(-3)}\frac{s_\sigma}{\kappa}v^\rho\right]\\&=s^\nu\frac{\partial T^{\mu\rho}_{(-2)}}{\partial x^\rho}\quad\text{from (47).}\end{aligned} \tag{48}$$

Employing an argument similar to that used in deriving (35) and (40) it can be shown that

$$\frac{\partial T^{\mu\rho}_{(-2)}}{\partial x^\rho}=-\frac{\partial T^{\mu\rho}_{(-3)}}{\partial x^\rho}=\left[\frac{\partial T^{\mu\rho}_{(-2)}}{\partial x^\rho}\right]_{(-3)}=-\left[\frac{\partial T^{\mu\rho}_{(-3)}}{\partial x^\rho}\right]_{(-3)}. \tag{49}$$

Therefore from (45), (48) and (49)

$$T'''^{\mu\nu} \equiv T''^{\mu\nu} - \theta\left[aT^{\mu\sigma}_{(-3)}\frac{s_\sigma}{\kappa}v^\nu\right] = T^{\mu\nu}_{-2)} + T^{\mu\sigma}_{(-3)}\frac{s_\sigma}{\kappa}v^\nu. \tag{50}$$

It is easy to see from (15), (16), (33) and (50) that

$$T'''^{\mu\nu} = T^{\mu\nu} + \theta K^{\mu\nu},$$

where $$\theta K^{\mu\nu} = \theta(a+b\phi+\dots)\,T^{\mu\nu} - \theta(a+b\phi+\dots)\,T^{\mu\nu}_{-2)} - \theta\left[aT^{\mu\sigma}_{(-3)}\frac{s_\sigma}{\kappa}v^\nu\right]$$

$$= \theta(a+b\phi+\dots)\,T^{\mu\nu}_{(-3} - \theta\left[aT^{\mu\sigma}_{(-3)}\frac{s_\sigma}{\kappa}v^\nu\right].$$

We can therefore write

$$K^{\mu\nu} = (a+b\phi+\dots)\,T^{\mu\nu}_{(-3} - aT^{\mu\sigma}_{(-3)}\frac{s_\sigma}{\kappa}v^\nu.$$

Using (39) one gets explicitly

$$K^{\mu\nu} = \left[T^{\mu\nu}_{(-3)} - T^{\mu\sigma}_{(-3)}\frac{s_\sigma}{\kappa}v^\nu\right] + \sum_{n=4}^{n=m}\left[\frac{T^{\mu\nu}_{(-n)}}{n-2} + \frac{T^{\mu\sigma}_{(-n)}s_\sigma/\kappa\,.\,v^\nu}{(n-2)(n-3)}\right]. \tag{51}$$

Thus $$T'''^{\mu\nu} = T^{\mu\nu}_{-2)} + T^{\mu\sigma}_{(-3)}\frac{s_\sigma}{\kappa}v^\nu = T^{\mu\nu} + \frac{\partial}{\partial x^\sigma}[K^{\mu\nu}s^\sigma - K^{\mu\sigma}s^\nu], \tag{52}$$

and is therefore of the required form.

Now $T'''^{\mu\nu}\kappa_\nu = T^{\mu\nu}_{-2)}\kappa_\nu + T^{\mu\sigma}_{(-3)}\frac{s_\sigma}{\kappa}\kappa'$ and therefore it possesses terms only up to the 2nd order, thus giving a finite integral over the world tube.

Since $$\frac{\partial}{\partial x^\rho} = \left(\delta^\sigma_\rho - \frac{s_\rho}{\kappa}v^\sigma\right)\frac{\partial}{\partial x^\sigma} + \frac{s_\rho}{\kappa}v^\sigma\frac{\partial}{\partial x^\sigma}, \tag{53}$$

it follows from (36) and (44) that

$$\left(\delta^\sigma_\rho - \frac{s_\rho}{\kappa}v^\sigma\right)\frac{\partial T^{\mu\nu}_{(-n)}}{\partial x^\sigma} = \left[\frac{\partial T^{\mu\nu}_{(-n)}}{\partial x^\rho}\right]_{(-n-1)}. \tag{54}$$

(44) and (54) hold not only for $T^{\mu\nu}_{(-n)}$ but also for any term of the nth order which is expressed as a/κ^n, where a is dependent on s^α/κ and τ only.

In view of (44) the conservation of $T'''^{\mu\nu}$ is easy to understand. Because, using (49),

$$\frac{\partial T'''^{\mu\nu}}{\partial x^\nu} = \left[\frac{\partial T^{\mu\nu}_{(-2)}}{\partial x^\nu}\right]_{(-3)} + \frac{\partial}{\partial x^\nu}\left[T^{\mu\sigma}_{(-3)}\frac{s_\sigma}{\kappa}v^\nu\right]$$

$$= \left[\frac{\partial T^{\mu\nu}_{(-2)}}{\partial x^\nu}\right]_{(-3)} + \frac{s_\sigma}{\kappa}v^\nu\frac{\partial T^{\mu\sigma}_{(-3)}}{\partial x^\nu} \quad \text{from (41)}$$

$$= \left[\frac{\partial T^{\mu\nu}_{(-2)}}{\partial x^\nu}\right]_{(-3)} + \left[\frac{\partial T^{\mu\nu}_{(-3)}}{\partial x^\nu}\right]_{(-3)} = 0 \quad \text{from (44) and (49).}$$

Precisely in the same way it can be proved that

$$T^{\mu\nu}_{(-n)}+T^{\mu\sigma}_{(-n-1)}\frac{s_\sigma}{\kappa}v^\nu \tag{55}$$

is conserved with respect to ν. Further it can be shown that if $n \geqslant 2$

$$T^{\mu\nu}_{(-n)}+T^{\mu\sigma}_{(-n-1)}\frac{s_\sigma}{\kappa}v^\nu = T^{\mu\nu}+\frac{\partial K^{\mu\nu\sigma}}{\partial x^\sigma},$$

where $K^{\mu\nu\sigma}$ is a tensor antisymmetrical in ν, σ. This follows from the fact that

$$\begin{aligned} S^{\mu\nu} &\equiv -T'''^{\mu\nu}+T^{\mu\nu}_{(-n)}+T^{\mu\sigma}_{(-n-1)}\frac{s_\sigma}{\kappa}v^\nu \\ &= \left(\delta^\nu_\sigma-\frac{s_\sigma}{\kappa}v^\nu\right)T^{\mu\sigma}_{(-3)}+T^{\mu\nu}_{(-4)}+\ldots+T^{\mu\nu}_{(-n)}+T^{\mu\sigma}_{(-n-1)}\frac{s_\sigma}{\kappa}v^\nu \end{aligned}$$

is conserved. It is then obvious that exactly as in the case of $T^{\mu\nu}$ a tensor

$$S'''^{\mu\nu} = S^{\mu\nu}_{(-2)}+S^{\mu\sigma}_{(-3)}\frac{s_\sigma}{\kappa}v^\nu = S^{\mu\nu}+\frac{\partial K'^{\mu\nu\sigma}}{\partial x^\sigma}$$

can be formed, where $K'^{\mu\nu\sigma}$ is a suitable tensor antisymmetrical in ν, σ. But since $S^{\mu\nu}$ does not contain terms of the 2nd order or lower, and

$$S^{\mu\rho}_{(-3)}\frac{s_\rho}{\kappa}v^\nu = v^\nu\frac{s_\rho}{\kappa}\left(\delta^\rho_\sigma-\frac{s_\sigma}{\kappa}v^\rho\right)T^{\mu\sigma}_{(-3)} = 0,$$

then $S'''^{\mu\nu} = 0$, i.e. $S^{\mu\nu} = -\dfrac{\partial K'^{\mu\nu\sigma}}{\partial x^\sigma}$, so that

$$T^{\mu\nu}_{(-n)}+T^{\mu\sigma}_{(-n-1)}\frac{s_\sigma}{\kappa}v^\nu = T'''^{\mu\nu}-\frac{\partial K'^{\mu\nu\sigma}}{\partial x^\sigma} = T^{\mu\nu}+\frac{\partial K^{\mu\nu\sigma}}{\partial x^\sigma}.$$

Thus it is found that there exist a large number of tensors formed according to the simple rule (55) and capable of being put in the form (1) in which the order k of the highest singularity can be any number such that $3 \leqslant k \leqslant m$.

4. It will now be proved that the 3rd order term in $T'''^{\mu\nu}$ does not give rise to a singularity in the energy-momentum integral, in the case of the electromagnetic and meson fields.

It has already been shown that $T'''^{\mu\nu}\kappa_\nu$ has terms only up to the order $1/\kappa^2$, even though $T'''^{\mu\nu}$ contains terms up to the order $1/\kappa^3$. If equation (4) is rewritten for $T'''^{\mu\nu}$, then

$$\frac{d}{d\tau}\left(\int_\kappa T'''^{\mu\nu}\frac{s_\nu}{\kappa}d\Omega\right) = -\frac{d}{d\kappa}\int_\kappa T'''^{\mu\nu}\kappa_\nu d\Omega.$$

Since the right side of this equation is finite the left side must also be so. It is, however, evident from (50) that the left side contains a singular term of the order $1/\kappa$ which is

$$\frac{d}{d\tau}\left(\int_{\kappa} T^{\mu\sigma}_{(-3)}\frac{s_\sigma}{\kappa}\,d\Omega\right). \tag{56}$$

Therefore this term must be put equal to zero. Therefore

$$\int_{\kappa} T^{\mu\sigma}_{(-3)}\frac{s_\sigma}{\kappa}\,d\Omega = \frac{c^\mu}{\kappa}, \tag{57}$$

where c^μ are constants. The physical meaning of this is obvious. Because the rate of outflow of energy and momentum from the particle to the outer space, given by $\int_{\kappa\to 0} T'''^{\mu\nu}\kappa_\nu\,d\Omega$ is finite, any singularity in total field energy or field momentum must remain independent of τ.*

It is now assumed that for values of τ very far in the past the world-line is straight so that the particle (whose motion the world-line represents) was moving a long time back in a straight line. It is further assumed that at that time it had no rotational motion. Consider now the following motion of the particle. In the remote past it was moving uniformly in the absence of an external field. It is then acted upon by an external field which causes acceleration. Finally, the external field dies down and the particle has again in the distant future uniform unaccelerated motion but with a velocity entirely different from the one it had far back in the past. If such a motion is permissible, then since c^μ is independent of τ, it must remain the same in the initial and final stages of motion, i.e. it should be the same whatever the velocity of the particle in the course of its uniform motion. This means that c^μ is independent of the choice of the Lorentz-frame of reference. The only tensors which are so independent are 0 and $g^{\mu\nu}$. The latter alternative is obviously inadmissible. Therefore $c^\mu = 0$.

---

* $\int_{\kappa=\epsilon}^{\kappa=\eta} d\kappa \int_{\kappa} T^{\mu\sigma}\frac{s_\sigma}{\kappa}\,d\Omega$ is not really the energy and momentum contained inside the volume enclosed between the two spheres $\kappa=\epsilon$ and $\kappa=\eta$. Energy-momentum is in fact defined as the integral of $T^{\mu\nu}$ over a plane three-dimensional space-like surface. However, if in equation (10) of A we take S as the plane defined by $[x^\mu - z^\mu(\tau)]\,v_\mu(\tau) = \epsilon$ (which is nothing but the space of the rest system taken a time ϵ later than τ), it is easy to see that P intersects S in the retarded sphere of radius ϵ. Remembering that for the plane $dS_\nu = v_\nu(\tau)\,dV$, where dV is the element of three-dimensional volume, relation (10) of A for $T'''^{\mu\nu}$ can be written as

$$\frac{d}{d\tau}\int_S T'''^{\mu\nu}v_\nu\,dV = \int_{\kappa=\epsilon} T'''^{\mu\nu}\kappa_\nu\,d\Omega - T'''^{\mu}(Q),$$

where $T'''^{\mu}(Q)$ is the integral over the two-dimensional surface of intersection of S and the tube Q. It can be directly inferred from this that the singularity in $\int T'''^{\mu\nu}v_\nu\,dV$ is independent of τ.

In the case of the electromagnetic and mesic particles it is usually assumed that such a motion is possible.* Therefore

$$c^\mu = 0 \tag{58}$$

for the corresponding fields.† Thus the integral of $T'''^{\mu\nu}$ over the light cone

$$\left(\int_{\kappa_1}^{\kappa_2} d\kappa \int_\kappa T'''^{\mu\nu} \frac{s_\nu}{\kappa} d\Omega\right)_{\kappa_1 \to 0}$$

is finite. From this it will be proved that the integral of $T'''^{\mu\nu}$ over any finite space-like surface is finite, irrespective of the manner in which the point of singularity is approached. One need consider only the 3rd order term $T^{\mu\sigma}_{(-3)} \frac{s_\sigma}{\kappa} v^\nu$, since this is the only one which can give rise to a singularity in the integral.

Let the space-like surface be

$$\zeta^0(x_\mu) = \text{constant}, \tag{59}$$

and let it intersect the world-line at c. Using the notation of A, the integral is

$$\int T^{\mu\sigma}_{(-3)} \frac{s_\sigma}{\kappa} v^\nu \frac{\partial \zeta^0}{\partial x^\nu} \,|\, D \,|^{-1} d\zeta^1 d\zeta^2 d\zeta^3. \tag{60}$$

Take $\zeta^1 = x^1$, $\zeta^2 = x^2$, $\zeta^3 = x^3$, where x^1, x^2, x^3 are measured in the rest system at c. Then $D = \partial\zeta^0/\partial x^0$. Now let c be surrounded by a two-dimensional closed surface lying in (59) and given by the equation

$$\zeta^1(x_\mu) = \text{constant} = \epsilon, \tag{61}$$

such that as $\epsilon \to 0$ this surface tends to the point c. It will be assumed that for sufficiently small values of ϵ (61) is always entirely enclosed between the two two-dimensional surfaces resulting from the intersection of the two tubes

$$\kappa = \alpha\epsilon, \tag{62}$$

$$\kappa = \beta\epsilon, \tag{63}$$

with (59). Here α and β are constants ($\alpha < \beta$) independent of ϵ. Now calculate the integral (60) taken over the portion of (59) lying between (61) and (63). For points lying in this region, it follows from the Taylor expansion of v^ν that for sufficiently small values of ϵ

$$|\, v^\nu - (v^\nu)_0 \,| < \lambda\epsilon,$$

where $(v^\nu)_0$ is the velocity at the point c and λ is a constant independent of ϵ. Therefore without affecting the singular terms v^ν can be replaced by $(v^\nu)_0$ in (60). Remembering that $(v^\nu)_0 = (1, 0, 0, 0)$ and $D = \partial\zeta^0/\partial x^0$ one gets the integral as

$$\int_{\zeta^1=\epsilon}^{\kappa=\beta\epsilon} T^{\mu\sigma}_{(-3)} \frac{s_\sigma}{\kappa} dx^1 dx^2 dx^3. \tag{64}$$

---

* In fact it may even be necessary to demand that only such motions are permissible (Dirac 1938).

† This can also be proved directly, from the form of the energy-momentum tensor of the electomagnetic and the meson fields.

Put
$$T^{\mu\sigma}_{(-3)}\frac{s_\sigma}{\kappa} = \frac{a^{\mu\sigma}_{(-3)}}{\kappa^3}\frac{s_\sigma}{\kappa} \equiv \frac{a^\mu}{\kappa^3},$$

so that a^μ is a finite function of τ and s^α/κ only. Let $|a^\mu|_m$ be the maximum value of $|a^\mu|$ in the region of integration of (64). Then

$$\left|\int_{\zeta^1=\epsilon}^{\kappa=\beta\epsilon}\frac{a^\mu}{\kappa^3}dx^1dx^2dx^3\right| < \int_{\zeta^1=\epsilon}^{\kappa=\beta\epsilon}\left|\frac{a^\mu}{\kappa^3}\right| dx^1dx^2dx^3 < \frac{|a^\mu|_m}{(\alpha\epsilon)^3}\int_{\kappa=\alpha\epsilon}^{\kappa=\beta\epsilon}dx^1dx^2dx^3. \qquad (65)$$

The spatial volume enclosed between $\kappa = \alpha\epsilon$ and $\kappa = \beta\epsilon$ is obviously of the order ϵ^3. The right side of (65) is thus finite and hence so also must be the left. Now it is seen from figure 1, where AB is the world-line, L the light cone and Q any finite tube, that since the integrals of $T'''^{\mu\nu}$ over P and L are non-singular the integral over the portion of S intercepted between P and Q must also be so. It has just now been shown that the integral taken over the portion of S enclosed between P and $\zeta^1 = \epsilon$ remains finite as $\epsilon \to 0$. Therefore the integral over the portion intercepted between $\zeta^1 = \epsilon$ and Q also remains finite as $\epsilon \to 0$. It has thus been proved that the integral over S is finite, no matter how the point of singularity be approached.*

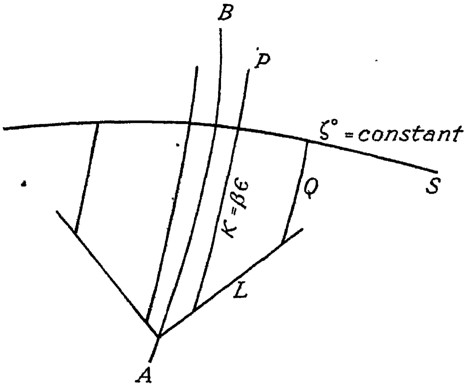


FIGURE 1

If $T'''^{\mu\nu}$ be integrated over the region of S enclosed between $\zeta^1 = \epsilon$ and $\zeta^1 = \eta$ $(\eta > \epsilon)$, then from (52) and equation (20) of A it is seen that

$$\int_S T'''^{\mu\nu}dS_\nu = \int_S T^{\mu\nu}dS_\nu + \int_{\zeta^1=\eta} K^{\mu\nu\sigma}dS_{\nu\sigma} - \int_{\zeta^1=\epsilon} K^{\mu\nu\sigma}dS_{\nu\sigma}, \qquad (66)$$

where $dS_{\nu\sigma}$ is written for

$$\frac{\partial\zeta^0}{\partial x^\nu}\frac{\partial\zeta^1}{\partial x^\sigma}|D|^{-1}d\zeta^2d\zeta^3$$

---

* It is noteworthy that the energy-momentum integrals though finite are not unique. Their value may depend on the way the point of singularity is approached. The energy is therefore indeterminate to this extent.

and $K^{\mu\nu\sigma}$ for $K^{\mu\nu}s^{\sigma}-K^{\mu\sigma}s^{\nu}$. It has been shown that as $\epsilon\rightarrow 0$ the left side remains finite. Now $K^{\mu\nu\sigma}$ contains certain terms of the 2nd order, namely, those arising from the 3rd order term $T^{\mu\nu}_{(-3)}-T^{\mu\sigma}_{(-3)}\frac{s_{\sigma}}{\kappa}v^{\nu}$ in $K^{\mu\nu}$. These terms can be removed and form a tensor $\tilde{T}^{\mu\nu}$ given by

$$\tilde{T}^{\mu\nu}=T^{\mu\nu}+\frac{\partial}{\partial x^{\sigma}}[K'^{\mu\nu}s^{\sigma}-K'^{\mu\sigma}s^{\nu}], \tag{67}$$

where

$$K'^{\mu\nu}=\sum_{n=4}^{n=m}\left[\frac{T^{\mu\nu}_{(-n}}{n-2}+\frac{T^{\mu\sigma}_{(-n)}\frac{s_{\sigma}}{\kappa}v^{\nu}}{(n-2)\,(n-3)}\right]. \tag{68}$$

Since only 2nd order terms have been removed from $K^{\mu\nu\sigma}$ it is easily seen from (66) that $\tilde{T}^{\mu\nu}$ also gives a finite integral as $\epsilon\rightarrow 0$. But as $K'^{\mu\nu\sigma}$ now contains no term of an order less than the 3rd its integral over the surface $\zeta^1=\eta$ tends to zero as $\eta\rightarrow\infty$,* so that

$$\int_S \tilde{T}^{\mu\nu}dS_{\nu}=\int_S T^{\mu\nu}dS_{\nu}-\int_{\zeta^1=\epsilon}K'^{\mu\nu\sigma}dS_{\nu\sigma}, \tag{69}$$

where the three-dimensional integrals are to be taken over the entire space outside $\zeta^1=\epsilon$. Since the total energy and momentum obtained by making ϵ tend to zero remain finite, the inflow must also be finite.

For small values of κ when $T^{\mu\nu}$ can be expanded into terms of different orders with respect to κ, it can be proved without difficulty that

$$\tilde{T}^{\mu\nu}=T^{\mu\nu}_{-3)}+s^{\nu}\frac{\partial T^{\mu\sigma}_{-3)}}{\partial x^{\sigma}}, \tag{70}$$

where $T^{\mu\nu}_{-3)}$ is the part of $T^{\mu\nu}$ left over after omitting terms of order higher than $1/\kappa^3$.

5. All the above results except (58) can be applied to any arbitrary tensor $U^{\alpha\beta\ldots\nu}$ having any number of indices, provided it satisfies a conservation equation with respect to ν, and can be expanded for small values of κ in a series involving different powers of κ. The tensor $U'''^{\alpha\ldots\nu}$, and therefore $\tilde{U}^{\alpha\ldots\nu}$ derived from $U^{\alpha\ldots\nu}$ by the above process, may, in general, still contain a singularity of the order $1/\kappa$ in the integral

$$\int_{\kappa}\tilde{U}^{\alpha\ldots\nu}\frac{s_{\nu}}{\kappa}d\Omega. \tag{71}$$

This singularity, however, must be independent of τ. It is now possible to show that this singularity can be removed.

It is easily seen by actual calculation that

$$\left\{\frac{\partial}{\partial x^{\sigma}}\left[\frac{(v^{\nu}s^{\sigma}-v^{\sigma}s^{\nu})}{\kappa^3}e^{-b\kappa}\log h\kappa\right]\right\}\frac{s_{\nu}}{\kappa}=\frac{1}{\kappa^3}(1-b\kappa\log h\kappa)\,e^{-b\kappa}, \tag{72}$$

---

* Here it is assumed that for sufficiently large values of η, $\xi^1=\eta$ is completely contained in the space between two spheres of radius $\alpha\eta$ and $\beta\eta$ described about c in the rest system. (α and β are constants independent of η.) The total surface of $\xi^1=\eta$ is assumed to tend to infinity as η^2.

where b and h are positive constants. Therefore

$$\int_\kappa \frac{\partial}{\partial x^\sigma}\left[\frac{(v^\nu s^\sigma - v^\sigma s^\nu)}{\kappa^3} e^{-b\kappa} \log h\kappa\right] \frac{s_\nu}{\kappa} d\Omega = \frac{4\pi}{\kappa}(1 - b\kappa \log h\kappa)\, e^{-b\kappa}. \tag{73}$$

The singularity in (71) is of the form $c^{\alpha\cdots}/\kappa$, where $c^{\alpha\cdots}$ are certain constants. Therefore

$$\int_\kappa \underline{U}^{\alpha\ldots\nu} \frac{s_\nu}{\kappa} d\Omega = bc^{\alpha\cdots} \log h\kappa\, e^{-b\kappa}, \tag{74a}$$

where
$$\underline{U}^{\alpha\ldots\nu} = \tilde{U}^{\alpha\ldots\nu} - \frac{\partial}{\partial x^\sigma}\left[\frac{c^{\alpha\cdots}}{4\pi} \frac{(v^\nu s^\sigma - v^\sigma s^\nu)}{\kappa^3} e^{-b\kappa} \log h\kappa\right]. \tag{74b}$$

Even though (74a) contains a singularity it can easily be seen that

$$\int_{\kappa=\epsilon}^{\kappa=\eta} d\kappa \int \underline{U}^{\alpha\ldots\nu} \frac{s_\nu}{\kappa} d\Omega$$

remains finite as $\epsilon \to 0$. Very much in the same manner as for $T'''^{\mu\nu}$ it can be shown that the integral of $\underline{U}^{\alpha\ldots\nu}$ over any space-like surface is finite no matter how the point of singularity be approached.

Further on account of $e^{-b\kappa}$ the surface integral of the term inside brackets in (74b) vanishes on the infinite sphere, so that in an equation like (69) the integral may still be omitted on the surface at infinity. It is to be noted that b is entirely arbitrary except for the fact that it is greater than zero.

Now apply these considerations to the case of the angular momentum tensor $M^{\mu\nu,\sigma}$ defined by

$$M^{\mu\nu,\sigma} = x^\mu T^{\nu\sigma} - x^\nu T^{\mu\sigma},$$

which is conserved with respect to σ, due to symmetry of the original tensor $T^{\mu\nu}$. One can therefore build $\tilde{M}^{\mu\nu,\sigma}$, which for small values of κ can be written as

$$\tilde{M}^{\mu\nu,\sigma} = M^{\mu\nu,\sigma}_{-3)} + s^\sigma \frac{\partial}{\partial x^\rho}(M^{\mu\nu,\rho}_{-3)}). \tag{75}$$

Since $x^\mu = z^\mu(\tau) + \frac{s^\mu}{\kappa}\kappa$ it is found that

$$M^{\mu\nu,\sigma}_{-3)} = x^\mu T^{\nu\sigma}_{-3)} - x^\nu T^{\mu\sigma}_{-3)} + s^\mu T^{\nu\sigma}_{(-4)} - s^\nu T^{\mu\sigma}_{(-4)}. \tag{76}$$

Therefore, using (57),

$$\int_\kappa \tilde{M}^{\mu\nu,\sigma}_{(-3)} \frac{s_\sigma}{\kappa} d\Omega = \frac{z^\mu c^\nu - z^\nu c^\mu}{\kappa} + \int (s^\mu T^{\nu\sigma}_{(-4)} - s^\nu T^{\mu\sigma}_{(-4)}) \frac{s_\sigma}{\kappa} d\Omega \equiv \frac{c^{\mu\nu}}{\kappa}.$$

Thus a tensor $\underline{M}^{\mu\nu,\sigma}$ can be constructed defined by

$$\underline{M}^{\mu\nu,\sigma} = M^{\mu\nu,\sigma} + \frac{\partial}{\partial x^\rho}[\Theta^{\mu\nu,\sigma} s^\rho - \Theta^{\mu\nu,\rho} s^\sigma], \tag{77a}$$

where $$\Theta^{\mu\nu,\sigma} = \sum_{n=4}^{n=m} \left[\frac{M^{\mu\nu,\sigma}_{(-n)}}{n-2} + \frac{M^{\mu\nu,\rho}_{(-n)} \frac{s_\rho}{\kappa} v^\sigma}{(n-2)\,(n-3)} \right] - \frac{c^{\mu\nu}}{4\pi} v^\sigma \frac{e^{-b\kappa} \log h\kappa}{\kappa^3} \tag{77b}$$

and $$c^{\mu\nu} = (z^\mu c^\nu - z^\nu c^\mu) + \int (s^\mu T^{\nu\sigma}_{(-4)} s_\sigma - s^\nu T^{\mu\sigma}_{(-4)} s_\sigma)\, d\Omega, \tag{77c}$$

such that the total angular momentum calculated from it is always finite.

Now return to the special case of the meson and the electromagnetic fields. Precisely as in the case of c^μ (equation (58)) it can be shown that since $c^{\mu\nu}$ is anti-symmetric in μ, ν while $g^{\mu\nu}$ is symmetric

$$c^{\mu\nu} = 0. \tag{78}$$

However, it is important to note that owing to the fact that the meson fields contain a factor $e^{-\chi\kappa}$ at infinity (Appendix, equation (116*a*)),* the angular-momentum integral as calculated from $M^{\mu\nu,\sigma}$ does not contain a singularity at infinity. In the electromagnetic case, however, a logarithmic singularity would appear at infinity if $c^{\mu\nu}$ did not vanish. However, as it is, it is found that the total angular momentum calculated from $\tilde{M}^{\mu\nu,\sigma}$ is free from singularity both at zero and infinity.†

All the above considerations are immediately applicable to an assembly of more than one particle, because near any particle the field due to all others is regular and can be included in the ingoing field. We have only to take instead of (67) and (68)

$$\tilde{T}^{\mu\nu} = T^{\mu\nu} + \sum_p \frac{\partial}{\partial x^\sigma} K^{\mu\nu\sigma}_p, \tag{78a}$$

where p refers to a particular particle,

$$K'^{\mu\nu\sigma}_p = K'^{\mu\nu}_p s^\sigma_p - K'^{\mu\sigma}_p s^\nu_p, \tag{78b}$$

$$K'^{\mu\nu}_p = \sum_{n=4}^{n=m} \left\{ \frac{T^{\mu\nu}_{p(-n)}}{n-2} + \frac{T^{\mu\sigma}_{p(-n)}}{(n-2)\,(n-3)} \frac{s_{\sigma,p}}{\kappa_p} v^\nu_p \right\}. \tag{78c}$$

---

* For large spatial distances from the particle (measured in the present space of any Lorentz-frame) the corresponding retarded point goes far back in the past of the particle, so that for sufficiently large distances the retarded point lies in the region of the world-line corresponding to uniform motion of the particle. The value of the fields at infinity may therefore be taken to be the same as in the static case. This is also evident from the consideration that since potentials and therefore fields are propagated with a velocity at most equal to that of light, the value of these quantities at sufficiently large distances must still be unaffected by radiation from the particle. The external field of course is supposed to vanish sufficiently fast at infinity.

† In this connexion it may be mentioned that a misprint has occurred in the expression for $F^{(2)}_{\mu\nu}$ given in the paper by Bhabha & Corben (1941). In equation (114) of their paper, instead of the term

$$\left[-2v_\mu s^\rho \frac{S_{\rho\nu}}{\kappa^3} \right]_-$$

it should read $$\left[-2v_\mu s^\rho \frac{\dot{S}_{\rho\nu}}{\kappa^3} \right]_-.$$

Here s_p^σ and v_p^ν refer to the particular particle. $T_{p(-n)}^{\mu\nu}$ is the nth order term in $T^{\mu\nu}$ when it is expanded in the neighbourhood of the particle p (i.e. for small κ_p). Similarly, corresponding to (77)

$$\tilde{M}^{\mu\nu,\sigma} = M^{\mu\nu,\sigma} + \sum_p \frac{\partial}{\partial x^\rho}[\Theta_p^{\mu\nu,\sigma} s_p^\rho - \Theta_p^{\mu\nu,\rho} s_p^\sigma], \tag{79a}$$

$$\Theta_p^{\mu\nu,\sigma} = \sum_{n=4}^{n=m_p}\left[\frac{M_{p(-n)}^{\mu\nu,\sigma}}{n-2} + \frac{M_{p(-n)}^{\mu\nu,\sigma}}{(n-2)(n-3)}\frac{s_{\rho,p}}{\kappa_p} v_p^\sigma\right] \tag{79b}$$

can be written. Here $c_p^{\mu\nu} = 0$.

6. Some special remarks may now be made regarding the symmetrization of the modified tensor for the electromagnetic dipole. As already observed in § 1 this symmetrization is, however, of no particular importance.

As proved in the appendix (equation (124)) it is found that for the electromagnetic case

$$T_{(-3)}^{\mu\nu\,\text{ret.}} s_\nu = 0, \tag{80}$$

where $T^{\mu\nu\,\text{ret.}}$ has the same meaning as that in the paper of Bhabha & Corben (1941). Therefore

$$T'''^{\mu\nu\,\text{ret.}} = T_{-2)}^{\mu\nu\,\text{ret.}} \quad \text{from (50)}, \tag{81}$$

which is obviously symmetrical. Since in the static case $T^{\mu\nu}$ does not contain terms of order lower than $1/\kappa^4$ (appendix, equation (116b))* then

$$T_{-2)\,\text{static}}^{\mu\nu} = 0. \tag{82}$$

Therefore the integral of $T_{-2)}^{\mu\nu\,\text{ret.}}$ does not contain any singularities at infinity.† Using the symbols used by Bhabha & Corben (1941), we write

$$T^{\mu\nu} = T^{\mu\nu\,\text{ret.}} + T^{\mu\nu\,\text{mix.}} + T^{\mu\nu\,\text{in.}}. \tag{83}$$

Because both $T^{\mu\nu\,\text{ret.}}$ and $T^{\mu\nu\,\text{in.}}$ are separately conserved $T^{\mu\nu\,\text{mix.}}$ must also be conserved, and the procedure of the preceding pages can be applied to $T^{\mu\nu\,\text{mix.}}$. We construct

$$T'''^{\mu\nu\,\text{mix.}} = T_{-2)}^{\mu\nu\,\text{mix.}} + T_{(-3)}^{\mu\sigma\,\text{mix.}}\frac{s_\sigma}{\kappa} v^\nu. \tag{84}$$

Further

$$\begin{aligned} M'''^{\mu\nu,\sigma\,\text{mix.}} = {} & (x^\mu T_{-2)}^{\nu\sigma\,\text{mix.}} - x^\nu T_{-2)}^{\mu\sigma\,\text{mix.}}) + (s^\mu T_{(-3)}^{\nu\sigma\,\text{mix.}} - s^\nu T_{(-3)}^{\mu\sigma\,\text{mix.}}) \\ & + (z^\mu T_{(-3)}^{\nu\rho\,\text{mix.}} - z^\nu T_{(-3)}^{\mu\rho\,\text{mix.}})\frac{s_\rho}{\kappa} v^\sigma + (s^\mu T_{(-4)}^{\nu\rho\,\text{mix.}} - s^\nu T_{(-4)}^{\mu\rho\,\text{mix.}})\frac{s_\rho}{\kappa} v^\sigma. \end{aligned} \tag{85}$$

From (84) and (85)

$$M'''^{\mu\nu,\sigma\,\text{mix.}} = x^\mu T'''^{\nu\sigma\,\text{mix.}} - x^\nu T'''^{\mu\sigma\,\text{mix.}} + \left[s^\mu\left\{T_{(-3)}^{\nu\rho\,\text{mix.}}\left(\delta_\rho^\sigma - \frac{s_\rho}{\kappa} v^\sigma\right) + T_{(-4)}^{\nu\rho\,\text{mix.}}\frac{s_\rho}{\kappa} v^\sigma\right\}\right]_-, \tag{86}$$

---

* The potentials contain terms of order $1/\kappa$ or higher. The fields therefore contain terms of order $1/\kappa^2$ or higher. Hence $T^{\mu\nu}$ does not contain terms of order lower than $1/\kappa^4$.

† Cf. footnote * on p. 159.

the minus sign at the end denoting the subtraction of terms got by exchange of μ, ν in the expression within the square brackets.

Since the highest singularity in the mixed tensor in the case of an electromagnetic dipole is of order $1/\kappa^3$, $T^{\nu\rho\,\text{mix.}}_{(-4)} = 0$. Using this fact and the conservation of $M'''^{\mu\nu,\sigma\,\text{mix.}}$ and $T'''^{\mu\sigma\,\text{mix.}}$, it is found from (86) that

$$0 = \frac{\partial M'''^{\mu\nu,\sigma\,\text{mix.}}}{\partial x^\sigma} = T'''^{\nu\mu\,\text{mix.}} - T'''^{\mu\nu\,\text{mix.}} + \frac{\partial}{\partial x^\sigma}\left[s^\mu T^{\nu\rho\,\text{mix.}}_{(-3)}\left(\delta^\sigma_\rho - \frac{s_\rho}{\kappa}v^\sigma\right)\right]_- . \tag{87}$$

Putting
$$L^{\mu\nu,\sigma} \equiv \left[s^\mu T^{\nu\rho\,\text{mix.}}_{(-3)}\left(\delta^\sigma_\rho - \frac{s_\rho}{\kappa}v^\sigma\right)\right]_- \tag{88}$$

it is found that
$$T'''^{\mu\nu\,\text{mix.}} - \frac{1}{2}\frac{\partial}{\partial x^\sigma}\left[L^{\mu\nu,\sigma} + L^{\sigma\mu,\nu} - L^{\nu\sigma,\mu}\right] \tag{89}$$

is symmetrical and is conserved (Pauli 1941). It is also in the form (1) as required. Further, since $L^{\mu\nu,\sigma}$ is of the order $1/\kappa^2$ the additional term does not introduce any singularity in the energy integral as is obvious from (66). Also, since $T^{\mu\nu\,\text{mix.}}_{(-3)}$ vanishes at infinity (due to the absence of the external field in the remote past),* the integral in (66) over the surface at infinity is zero.

Thus
$$\overline{T}^{\mu\nu} = T^{\mu\nu\,\text{ret.}}_{-2)} + T'''^{\mu\nu\,\text{mix.}} - \frac{1}{2}\frac{\partial}{\partial x^\sigma}\left[L^{\mu\nu,\sigma} + L^{\sigma\mu,\nu} - L^{\nu\sigma,\mu}\right] + T^{\mu\nu\,\text{in.}}$$

can be taken as the modified tensor. This tensor is symmetrical. That in the case of a point charge this symmetry is possible has already been shown by Pryce (1938). In this case

$$T^{\mu\nu\,\text{ret.}} = \frac{e^2}{4\pi}\left[-\frac{(1-\kappa')^2}{\kappa^6}s^\mu s^\nu + (1-\kappa')\frac{(v^\mu s^\nu + v^\nu s^\mu)}{\kappa^5} + \frac{(\dot{v}^\mu s^\nu + \dot{v}^\nu s^\mu) - (\dot{v})^2 s^\mu s^\nu - \frac{1}{2}g^{\mu\nu}}{\kappa^4}\right]. \tag{90}$$

Thus according to (51)

$$K^{\mu\nu} = \left(T^{\mu\nu}_{(-3)} - T^{\mu\sigma}_{(-3)}\frac{s_\sigma}{\kappa}v^\nu\right) + \frac{T^{\mu\nu}_{(-4)}}{2} + \frac{T^{\mu\sigma}_{(-4)}}{2\,.\,1}\frac{s_\sigma}{\kappa}v^\nu.$$

---

* For small κ the retarded potentials U_ν and the retarded field quantities $G_{\mu\nu}$ can be expanded in a series of positive and negative powers of κ, and the ingoing field in a Taylor's series with positive powers of κ

$$G^{\text{in.}}_{\mu\nu} = (G^{\text{in.}}_{\mu\nu})_{z_\rho(\tau_0)} + \left(\frac{\partial G^{\text{in.}}_{\mu\nu}}{\partial x^\sigma}\right)_{z_\rho(\tau_0)}\frac{s^\sigma}{\kappa}\kappa + \frac{1}{2}\left(\frac{\partial^2 G^{\text{in.}}_{\mu\nu}}{\partial x^\sigma \partial x^\lambda}\right)\frac{s^\sigma}{\kappa}\frac{s^\lambda}{\kappa}\kappa^2 + \dots,$$

where the suffix $z_\rho(\tau_0)$ means that the corresponding quantities are to be evaluated at the retarded point. Therefore $T^{\mu\nu\,\text{mix.}}_{(-3)}$ consists of terms of the form $\frac{u\,.\,v}{\kappa^3}$, where u and v are both of order zero in κ and u comes from $G^{\text{ret.}}_{\mu\nu}$, while v comes from $G^{\text{in.}}_{\mu\nu}$. The factor in v having the suffix $z_\rho(\tau_0)$ is to be evaluated at the retarded point. At infinity $v = 0$ because the corresponding retarded point goes to the remote past and there the field quantities and their differentials vanish.

(In spite of the fact that $K^{\mu\nu}$ is taken instead of $K'^{\mu\nu}$ given by (68), the third term in (66) vanishes as $\eta \to \infty$ due to (82).*) So that

$$K^{\mu\nu\sigma} = K^{\mu\nu}s^{\sigma} - K^{\mu\sigma}s^{\nu}$$

$$= \frac{e^2}{4\pi}\left[(\tfrac{3}{4}-\kappa')\frac{s^{\mu}}{\kappa^5}(v^{\nu}s^{\sigma}-v^{\sigma}s^{\nu}) + s^{\mu}\frac{(\dot{v}^{\nu}s^{\sigma}-\dot{v}^{\sigma}s^{\nu})}{\kappa^4} - \frac{1}{4}\frac{(g^{\mu\nu}s^{\sigma}-g^{\mu\sigma}s^{\nu})}{\kappa^4}\right]. \tag{91}$$

It is interesting to see that this is different from Pryce's tensor which is more complicated. It is (Pryce 1938)

$$K^{\mu\nu,\sigma} = \frac{e^2}{4\pi}\left[-\tfrac{9}{4}\kappa'\frac{(v^{\nu}s^{\sigma}-v^{\sigma}s^{\nu})}{\kappa^5}s^{\mu} + \frac{(g^{\mu\nu}v^{\sigma}-g^{\mu\sigma}v^{\nu})}{4\kappa^3} - (1+2\kappa')\frac{(g^{\mu\nu}s^{\sigma}-g^{\mu\sigma}s^{\nu})}{4\kappa^4}\right.$$
$$\left. + \frac{3(v^{\nu}s^{\sigma}-v^{\sigma}s^{\nu})}{4\kappa^4}v^{\mu} + \frac{3(\dot{v}^{\nu}s^{\sigma}-\dot{v}^{\sigma}s^{\nu})}{4\kappa^4}s^{\mu}\right].$$

The modified tensor obtained from both these expressions is symmetric. Our modified tensor is just equal to $T^{\mu\nu\,\text{ret.}}_{-2)}$ as is proved quite generally in (81).

In conclusion I wish to express my thanks to Professor H. J. Bhabha for valuable discussions, general guidance and suggested improvements in the writing of this paper.

APPENDIX

Consider now
$$\kappa\frac{d}{d\kappa}\int T^{\mu\nu}\kappa_{\nu}\,.\,\kappa^2 d\omega, \tag{92}$$

τ being kept constant during differentiation. Here $d\omega$ is an element of solid angle in the rest system and $d\Omega = \kappa^2 d\omega$. Since

$$\kappa = s_{\mu}v^{\mu} \equiv [x_{\mu} - z_{\mu}(\tau)]\,v^{\mu}(\tau),$$

on differentiation keeping τ constant

$$d\kappa = dx_{\mu}v^{\mu}. \tag{93a}$$

For constant τ
$$s^{\mu}dx_{\mu} = 0. \tag{93b}$$

These are the only two conditions which the infinitesimal vector which connects a point on the sphere κ to a corresponding point on the slightly larger sphere $\kappa + d\kappa$, need satisfy. The way in which the points on these two spheres may be put in correspondence is immaterial. Now take dx_{μ} along s_{μ} thus satisfying (93b) automatically. (93a) then gives

$$ds_{\mu} = dx_{\mu} = d\kappa\frac{s_{\mu}}{\kappa},$$

i.e.
$$\frac{ds_{\mu}}{d\kappa} = \frac{s_{\mu}}{\kappa}. \tag{94}$$

---

* Cf. footnote * on p. 159.

From (94) it is easily found that

$$\frac{d}{d\kappa}\left(\frac{s_\mu}{\kappa}\right) = 0, \tag{95}$$

from which it may be concluded that

$$\frac{d}{d\kappa}(d\omega) = 0 \tag{96}$$

as $d\omega$ is dependent on s_μ/κ only. Thus

$$\frac{d}{d\kappa}\int T^{\mu\nu}\kappa_\nu\,.\,\kappa^2 d\omega = \int\frac{\partial}{\partial x^\sigma}(T^{\mu\nu}\kappa_\nu)\frac{dx^\sigma}{d\kappa}\kappa^2 d\omega + 2\int T^{\mu\nu}\kappa_\nu\kappa\,d\omega.$$

Since $\dfrac{\partial\kappa_\nu}{\partial x^\sigma}\dfrac{dx^\sigma}{d\kappa} = \dfrac{\kappa'}{\kappa}\dfrac{s_\nu}{\kappa}$ (from equations (12) of A, (94) and (95))

then

$$\kappa\frac{d}{d\kappa}\int T^{\mu\nu}\kappa_\nu d\Omega = \int\left[2T^{\mu\nu}+s^\sigma\frac{\partial T^{\mu\nu}}{\partial x^\sigma}\right]\kappa_\nu d\Omega + \int T^{\mu\nu}\frac{s_\nu}{\kappa}\kappa'\,d\Omega$$

$$= \int\frac{\partial}{\partial x^\sigma}[T^{\mu\nu}s^\sigma - T^{\mu\sigma}s^\nu]\,.\,\kappa_\nu d\Omega, \tag{97}$$

as $\dfrac{\partial}{\partial x^\sigma}[T^{\mu\nu}s^\sigma - T^{\mu\sigma}s^\nu] = 2T^{\mu\nu}+s^\sigma\dfrac{\partial T^{\mu\nu}}{\partial x^\sigma} + T^{\mu\sigma}\dfrac{s_\sigma}{\kappa}v^\nu,$ (cf. equation (6*b*))

if $\dfrac{\partial T^{\mu\nu}}{\partial x^\nu} = 0$. Since $\dfrac{d}{d\kappa}\left(\dfrac{s_\mu}{\kappa}\right) = 0$ it is easy to prove further that

$$\kappa\frac{d}{d\kappa}\int T^{\mu\nu}\kappa_\nu\frac{s^\alpha}{\kappa}\cdots\frac{s^\lambda}{\kappa}d\Omega = \int\frac{\partial}{\partial x^\sigma}[T^{\mu\nu}s^\sigma - T^{\mu\sigma}s^\nu]\,.\,\kappa_\nu\frac{s^\alpha}{\kappa}\cdots\frac{s^\lambda}{\kappa}d\Omega, \tag{98}$$

as quoted in the text. Remembering (94) and (95) it can similarly be proved that

$$\frac{d}{d\kappa}\int T^{\mu\nu}s_\nu\frac{s^\alpha}{\kappa}\cdots\frac{s^\lambda}{\kappa}d\Omega = \frac{1}{\kappa}\int\left[3T^{\mu\nu}+s^\sigma\frac{\partial T^{\mu\nu}}{\partial x^\sigma}\right]s_\nu\frac{s^\alpha}{\kappa}\cdots\frac{s^\lambda}{\kappa}d\Omega, \tag{99}$$

so that

$$\kappa\frac{d}{d\kappa}\int T^{\mu\nu}s_\nu\frac{s^\alpha}{\kappa}\cdots\frac{s^\lambda}{\kappa}d\Omega = \int\left[2T^{\mu\nu}+s^\sigma\frac{\partial T^{\mu\nu}}{\partial x^\sigma}+T^{\mu\sigma}\frac{s_\sigma}{\kappa}v^\nu\right]s_\nu\frac{s^\alpha}{\kappa}\cdots\frac{s^\lambda}{\kappa}d\Omega$$

$$= \int\theta T^{\mu\nu}\,.\,s_\nu\frac{s^\alpha}{\kappa}\cdots\frac{s^\lambda}{\kappa}d\Omega. \tag{100}$$

In the same way it is found that

$$\frac{d}{d\kappa}\int T^{\mu\sigma}(\delta^\nu_\sigma - v_\sigma v^\nu)\frac{s^\alpha}{\kappa}\cdots\frac{s^\lambda}{\kappa}d\Omega = \frac{1}{\kappa}\int\left[2T^{\mu\sigma}+s^\rho\frac{\partial T^{\mu\sigma}}{\partial x^\rho}\right](\delta^\nu_\sigma - v_\sigma v^\nu)\frac{s^\alpha}{\kappa}\cdots\frac{s^\lambda}{\kappa}d\Omega, \tag{101}$$

because $\delta^\nu_\sigma - v_\sigma v^\nu$, which depends on τ alone, is to be treated as a constant here. Since $v^\sigma(\delta^\nu_\sigma - v_\sigma v^\nu) = 0$ then

$$\kappa\frac{d}{d\kappa}\int T^{\mu\sigma}(\delta^\nu_\sigma - v_\sigma v^\nu)\frac{s^\alpha}{\kappa}\cdots\frac{s^\lambda}{\kappa}d\Omega = \int\theta T^{\mu\sigma}\,.\,(\delta^\nu_\sigma - v_\sigma v^\nu)\frac{s^\alpha}{\kappa}\cdots\frac{s^\lambda}{\kappa}d\Omega. \tag{102}$$

The most general equation which the potentials U_ν of a meson field may have to satisfy is

$$\frac{\partial}{\partial x^\rho}\frac{\partial}{\partial x_\rho}U^\nu + \chi^2 U^\nu = 4\pi\frac{\partial}{\partial x^\alpha}\cdots\frac{\partial}{\partial x^\lambda}\Sigma^{\alpha\ldots\lambda,\nu}, \tag{103}$$

where $\Sigma^{\alpha\ldots\lambda,\nu}$ is a tensor describing the effect of a multipole on the field. In fact, in the most general case the right side is a sum of terms containing different numbers of differentiations, but it will be sufficient to consider one such term since the solution of this general case is just the sum of the solutions of each of the equations involving one term only. Since interest arises only in the case of point-particles we can write (cf. Bhabha 1941, p. 321)

$$\Sigma^{\alpha\ldots\lambda,\nu}_{(x_\rho)} = \int_{-\infty}^{\infty} d\tau\, S^{\alpha\ldots\lambda,\nu}_{(\tau)}\,\delta(x^0-z^0)\,\delta(x^1-z^1)\,\delta(x^2-z^2)\,\delta(x^3-z^3). \tag{104}$$

As is well known the solution for the retarded potentials is

$$U^{\nu\,\text{ret.}}_{(x_\rho)} = \frac{\partial}{\partial x^\alpha}\cdots\frac{\partial}{\partial x^\lambda}\left(\frac{S^{\alpha\ldots\lambda,\nu}}{\kappa}\right) - \chi\frac{\partial}{\partial x^\alpha}\cdots\frac{\partial}{\partial x^\lambda}\int_{-\infty}^{\tau_0} S^{\alpha\ldots\lambda,\nu}\frac{J_1(\chi u)}{u}\,d\tau, \tag{105}$$

where $u_\rho(\tau) = x_\rho - z_\rho(\tau)$, $u = (u_\rho u^\rho)^{\frac{1}{2}}$ and τ_0 is the retarded time given by $u(\tau_0) = 0$. Here the first term is just the one which alone appears for the electromagnetic field ($\chi = 0$). It is obviously expressible in a finite series involving terms of different orders in κ. The differentiations in the 2nd term act in two ways. First, on the limit of integration τ_0, and secondly, on the integrand. Bearing in mind that

$$\frac{\partial}{\partial x^\lambda}\left(\frac{J_n(\chi u)}{u^n}\right) = -\chi u_\lambda\frac{J_{n+1}(u)}{u^{n+1}} \tag{106}$$

and

$$\frac{\partial u_\lambda}{\partial x^\alpha} = g_{\lambda\alpha}, \tag{107}$$

then

$$\frac{\partial}{\partial x^\nu}\cdots\frac{\partial}{\partial x^\lambda}\left(\frac{J_1(\chi u)}{u}\right) \tag{108}$$

can be evaluated in terms of u^α and $\frac{J_n(\chi u)}{u^n}$. Also, since

$$u^\alpha(\tau_0) = s^\alpha \tag{109}$$

and

$$\left(\frac{J_n(\chi u)}{u^n}\right)_{\tau_0} = \left(\frac{J_n(\chi u)}{u^n}\right)_{u\to 0} = \frac{\chi^n}{2^n\,.\,n!}, \tag{110}$$

then the value of (108) at the retarded point can be expressed as a finite series in s^α involving terms of different orders in κ. It should be noted that

$$\frac{\partial}{\partial x^\mu}\int_{-\infty}^{\tau_0} S^{\alpha\ldots,\gamma}\frac{\partial}{\partial x^\nu}\cdots\frac{\partial}{\partial x^\lambda}\left(\frac{J_1(\chi u)}{u}\right)d\tau$$

$$= \frac{s_\mu}{\kappa}\left[S^{\alpha\ldots,\gamma}\frac{\partial}{\partial x^\nu}\cdots\frac{\partial}{\partial x^\lambda}\left(\frac{J_1(\chi u)}{u}\right)\right]_{\tau_0} + \int_{-\infty}^{\tau_0} S^{\alpha\ldots,\gamma}\frac{\partial}{\partial x^\mu}\frac{\partial}{\partial x^\nu}\cdots\frac{\partial}{\partial x^\lambda}\left(\frac{J_1(\chi u)}{u}\right)d\tau. \tag{111}$$

Again the first term on the right and its differentials can be written as finite series in s^α containing terms of different orders in κ. Therefore the only term in $U^{\nu\,\mathrm{ret.}}$ which has to be considered is

$$\int_{-\infty}^{\tau_0} S^{\alpha\ldots\lambda,\nu}\frac{\partial}{\partial x^\alpha}\cdots\frac{\partial}{\partial x^\lambda}\left(\frac{J_1(\chi u)}{u}\right)d\tau. \qquad (112)$$

As shown above this will consist of terms of the type

$$\int_{-\infty}^{\tau_0} S^{\alpha\ldots\lambda,\nu}u_\gamma\ldots u_\mu\frac{J_n(\chi u)}{u^n}\,d\tau. \qquad (113)$$

It is noteworthy that all these terms are non-singular. They can be expanded in a series containing only terms of positive order in κ in the following way. Write $u_\rho \equiv x_\rho - z_\rho(\tau) = l_\rho - s_\rho$, where $l_\rho \equiv z_\rho(\tau) - z_\rho(\tau_0)$, so that $u^2 = l_\rho l^\rho - 2l^\rho s_\rho = l^2 - 2(ls)$, where $l = (l_\rho l^\rho)^{\frac{1}{2}}$ and $(ls) = (l_\rho s^\rho)$. Then

$$\frac{J_n(\chi u)}{u^n} = \frac{J_n(\chi\sqrt{[l^2-2(ls)]})}{(\sqrt{[l^2-2(ls)]})^n}. \qquad (114)$$

Since $\dfrac{J_n(\chi l)}{l^n}$ is a continuous and differentiable function of l^2, (114) can be written in the form of a Taylor's series for sufficiently small values of s_ρ

$$\begin{aligned}\frac{J_n(\chi u)}{u^n} &= \frac{J_n(\chi l)}{l^n} + \sum_{p=1}^{\infty}\frac{[-2(ls)]^p}{p!}\left(\frac{d}{dl^2}\right)^p\frac{J_n(\chi l)}{l^n}\\ &= \frac{J_n(\chi l)}{l^n} + \sum_{p=1}^{\infty}\chi^p\frac{(ls)^p}{p!}\frac{J_{n+p}(\chi l)}{l^{n+p}}. \qquad (115)\end{aligned}$$

It follows from (113) and (115) that the expansion of (105) in a series of s^α is possible. s^α are independent of τ and so can be put outside the integral sign. Since the integrands involve only $S^{\alpha\ldots\lambda,\nu}$ and l_λ which are functions of τ and τ_0 alone, the integral is a function of τ_0 alone. It is now possible to separate terms of different orders by writing s^α/κ instead of s^α and multiplying by a suitable power of κ to compensate it. Thus the expansion of $U^{\nu\,\mathrm{ret.}}$ in the required form is obtained. $G^{\mathrm{ret.}}_{\mu\nu}$, which involves one more differentiation, can be expanded in an entirely similar manner. Since the ingoing field and potentials are always expressible as such a series by Taylor's expansion, $T^{\mu\nu}$ can be written in the required form with different order terms separated. This result is used in (9) and (10) in the text.

As is well known, in the static case when $S^{\alpha\ldots\lambda,\nu}$ is constant we get

$$U^\nu = S^{ik\ldots m,\nu}\frac{\partial}{\partial x^i}\frac{\partial}{\partial x^k}\cdots\frac{\partial}{\partial x^m}\left(\frac{e^{-\chi\kappa}}{\kappa}\right). \qquad (116a)$$

(Here the latin indices run from 1 to 3 only.) In the electromagnetic case, by putting $\chi = 0$, we get

$$U^\nu = S^{ik\ldots m,\nu}\frac{\partial}{\partial x^i}\frac{\partial}{\partial x^k}\cdots\frac{\partial}{\partial x^m}\left(\frac{1}{\kappa}\right). \qquad (116b)$$

From (105) it is found that in the electromagnetic case the retarded potentials* ϕ^ν do not contain terms of order lower than $1/\kappa$. These satisfy the equation

$$\frac{\partial \phi^\nu}{\partial x^\nu} = 0. \tag{117}$$

Using the relations (44) and (54), it is found from (117) that

$$\frac{s_\nu}{\kappa} v^\sigma \frac{\partial \phi^\nu_{(-n)}}{\partial x^\sigma} + \left(\delta^\sigma_\nu - \frac{s_\nu}{\kappa} v^\sigma\right) \frac{\partial \phi^\nu_{(-n+1)}}{\partial x^\sigma} = 0 \quad \text{for} \quad n > 1 \tag{118a}$$

and

$$\frac{s_\nu}{\kappa} v^\sigma \frac{\partial \phi^\nu_{(-1)}}{\partial x^\sigma} = 0. \tag{118b}$$

Here $(-n)$ denotes the nth order term. Now calculate

$$[F_{\mu\nu(-n)} F^{\nu\rho}_{(-1)} + F_{\mu\nu(-1)} F^{\nu\rho}_{(-n)} + \tfrac{1}{2}\delta^\rho_\mu F_{\alpha\beta(-1)} F^{\alpha\beta}_{(-1)}]\, s_\rho, \tag{119}$$

where $F_{\mu\nu}$ are the retarded electromagnetic field strengths. From the relation

$$F_{\mu\nu} = \frac{\partial \phi_\nu}{\partial x^\mu} - \frac{\partial \phi_\mu}{\partial x^\nu} \tag{120}$$

it is found that

$$F_{\mu\nu(-n)} = \left[\frac{s_\mu}{\kappa} v^\sigma \frac{\partial \phi_{\nu(-n)}}{\partial x^\sigma} + \left(\delta^\sigma_\mu - \frac{s_\mu}{\kappa} v^\sigma\right) \frac{\partial \phi_{\nu(-n+1)}}{\partial x^\sigma}\right]_- \quad \text{for} \quad n > 1 \tag{121a}$$

and

$$F_{\mu\nu(-1)} = \left[\frac{s_\mu}{\kappa} v^\sigma \frac{\partial \phi_{\nu(-1)}}{\partial x^\sigma}\right]_- . \tag{121b}$$

Using (121b) and (118b) it is found that $F_{\mu\nu(-1)}\, s^\nu = 0$, so that the first term in (119) does not contribute anything. Now

$$\begin{aligned} F_{\mu\nu(-n)}\, s^\nu &= \left[\frac{s_\mu}{\kappa} s^\nu v^\sigma \frac{\partial \phi_{\nu(-n)}}{\partial x^\sigma} + s^\nu\left(\delta^\sigma_\mu - \frac{s_\mu}{\kappa} v^\sigma\right) \frac{\partial \phi_{\nu(-n+1)}}{\partial x^\sigma} - s^\sigma \frac{\partial \phi_{\mu(-n+1)}}{\partial x^\sigma}\right] \\ &= \left[(s^\nu \delta^\sigma_\mu - s_\mu g^{\sigma\nu}) \frac{\partial \phi_{\nu(-n+1)}}{\partial x^\sigma} - s^\sigma \frac{\partial \phi_{\mu(-n+1)}}{\partial x^\sigma}\right] \quad \text{from (118a)}, \end{aligned} \tag{122}$$

so that

$$F_{\mu\nu(-1)} F^{\nu\rho}_{(-n)}\, s_\rho = F_{\mu\nu(-1)} \left\{ s^\rho \frac{\partial \phi_{\rho(-n+1)}}{\partial x^\nu} - s^\rho \frac{\partial \phi^\nu{}_{(-n+1)}}{\partial x^\rho} \right\} \tag{123}$$

as $F_{\mu\nu(-1)}\, s^\nu = 0$. Also

$$\begin{aligned} F_{\alpha\beta(-n)} F^{\alpha\beta}_{(-1)} &= 2F^{\alpha\beta}_{(-1)} \left[\frac{s_\alpha}{\kappa} v^\sigma \frac{\partial \phi_{\beta(-n)}}{\partial x^\sigma} + \left(\delta^\sigma_\alpha - \frac{s_\alpha}{\kappa} v^\sigma\right) \frac{\partial \phi_{\beta(-n+1)}}{\partial x^\sigma}\right] \\ &= 2F^{\alpha\beta}_{(-1)} \frac{\partial \phi_{\beta(-n+1)}}{\partial x^\alpha}. \end{aligned}$$

---

* ϕ^ν is used to denote $U^{\nu\,\text{ret.}}$ for the electromagnetic field.

Thus (119) becomes

$$\left[\frac{s_\mu}{\kappa} v^\sigma \frac{\partial \phi_{\nu(-1)}}{\partial x^\sigma}\right]\left[s^\rho \frac{\partial \phi_{\rho(-n+1)}}{\partial x_\nu} - s^\rho \frac{\partial \phi^\nu_{(-n+1)}}{\partial x_\rho}\right] + s_\mu \left[\frac{s^\alpha}{\kappa} v^\sigma \frac{\partial \phi^\beta_{(-1)}}{\partial x^\sigma}\right]_- \frac{\partial \phi_{\beta(-n+1)}}{\partial x^\alpha} = 0.$$

Thus the expression (119) vanishes. Since $4\pi T^{\mu\rho\,\text{ret.}}_{(-3)}\, s_\rho$ is just this expression with $n = 2$, then

$$T^{\mu\nu\,\text{ret.}}_{(-3)}\, s_\nu = 0. \tag{124}$$

REFERENCES

Bhabha 1939 *Proc. Roy. Soc.* A, **172**, 384–409.
Bhabha 1941 *Proc. Roy. Soc.* A, **178**, 315–350.
Bhabha & Corben 1941 *Proc. Roy. Soc.* A, **178**, 273–314.
Bhabha & Harish-Chandra 1944 *Proc. Roy. Soc.* A, **183**, 134.
Dirac 1938 *Proc. Roy. Soc.* A, **167**, 148–169.
Pauli 1941 *Rev. Mod. Phys.* **13**, 203–232.
Pryce 1938 *Proc. Roy. Soc.* A, **168**, 389–401.

The combination of fibrous proteins with acids

II. The adsorption of dye anions

BY G. A. GILBERT

(*Communicated by E. K. Rideal, F.R.S.—Received* 7 *February* 1944)

Equations developed previously for the titration of fibrous proteins with acids are shown to accord with extant data for the adsorption of dyes in the presence of salt. The contribution of particular groups in the dye molecules to the free energy of adsorption is determined, and the influence of the valency of the dye ion is related to the potential on the protein fibre. The effect of temperature on the ionic equilibria is calculated, and estimated experimentally for a simple dye using a rapid 'null' method for the attainment of equilibrium.

INTRODUCTION

In Part I (Gilbert & Rideal 1944, quoted as (I)) the adsorption of pure acids by fibrous proteins was examined. It was shown that equal numbers of anions and cations were adsorbed by an initially isoelectric fibre, the fibre as a whole becoming positively or negatively charged to achieve this equivalence.

I. FACTORS INFLUENCING THE AFFINITY OF DYES

Ions are bound to proteins by electrostatic forces arising from the net charge of the protein, and by more specific forces of shorter range involving the immediate environment of the ions. With this in mind, consider wool keratin in equilibrium with dilute sulphuric acid, sodium sulphate and the sodium salt of a dye acid such as Orange 2G. If the dye is fairly simple, it can be stated in broad outline how the equilibria will be affected by changes in the relative amounts of the constituents (cf. Hallitt 1899, Bancroft 1914, 1915).

Protons and dye anions are strongly adsorbed through their high affinity for keratin. If equivalently there are more protons present than dye ions, sulphate ions are adsorbed appreciably in spite of their low intrinsic affinity; if there are less protons than dye ions, sodium ions are adsorbed.

Assuming that equilibrium can be reached, the final distribution of ions is determined by their intrinsic affinities and concentrations. A model was used in (I) to express this quantitatively, and involved assigning to the fibre an equal number of positive and negative sites on which anions and cations respectively were randomly distributed. No allowance was made for the adsorption of more than one ion on the same site, nor for the adsorption of one kind of ion by more than one type of site. That there are circumstances when more than one type of site comes into play is obvious from the titration curves of keratin titrated with weak acids. It is probable that charged molecules with a very high intrinsic affinity for keratin (even when they contain such strongly ionized groups as $—SO_3'$) can be adsorbed by other than positively charged sites, and then perhaps form sites themselves for the adsorption of cations. In addition, the more complex the anion adsorbed, the less is it likely to be legitimate to treat all the positive sites as identical. However, if the simple model fails, it will be difficult to determine what approximations to apply, for other factors come into play. Thus molecules as large as dye anions probably find some sites extremely difficult to reach at temperatures which do not cause fibre decomposition, and may furthermore undergo aggregation through the very forces that confer their high affinity.

For the ideal system represented by the model, it was shown in (I) that the fraction θ_i of the total sites occupied by an ion of type i and valency Z was given by (cf. equation (31) (I))

$$RT\ln\frac{\theta_i}{1-\sum_i \theta_i} = RT\ln C_i - \Delta\mu_i - ZF\psi, \tag{1}$$

if activity coefficients in solution are neglected. R is the gas constant, T the absolute temperature, C_i the concentration of the ion in the solution, F the Faraday, ψ the fibre potential, and $\Delta\mu_i$ the intrinsic affinity. (I) should be consulted for the precise definition of the symbols.

The term $\left(\frac{\theta_i}{1-\sum_i \theta_i}\right)$ plays a similar role in the fibre to that of the concentration C_i in solution, and is increased by decreasing the adsorption of other kinds of ions of the same sign, or by reducing $\Delta\mu_i$ or $Z\psi$. This can be illustrated by some experiments of Speakman & Clegg (1934).

The influence of the valency of a dye ion

First, find how the potential ψ will vary if sodium sulphate is increased in concentration in the system wool keratin, H_2SO_4, Na_2SO_4, Na_ZD (where D is a dye anion of valency Z). To correspond to the conditions of Speakman & Clegg's experiments, let it be supposed that the sulphuric acid present is insufficient to saturate the fibre, but is still in fair excess compared with the dye, and let the volume of the solution per unit weight of fibre be small enough for most of the protons present to be adsorbed. They will be accompanied by the majority of the dye ions, and also by sulphate ions almost sufficient in number to neutralize the change resulting from the protons and dye ions. When the salt concentration is low adsorbed sodium ions can be neglected, since under the conditions above the fibre is positively charged, and the affinity of the sodium ions is reduced to a very low value indeed by electrostatic repulsion.

Rewriting equation (1) with $Z = 2$ for the sulphate ion, replacing C_i by SO_4 (the external sulphate ion concentration), and allowing for the ion being negative,

$$\psi = \frac{RT}{2F}\ln\left(\frac{\theta_{SO_4}}{1-\sum_i \theta_i}\right)+\frac{\Delta\mu_{SO_4}}{2F}-\frac{RT}{2F}\ln SO_4 \tag{2}$$

is obtained for the fibre potential. The expansion of $(1-\sum_i \theta_i)$ is $(1-\theta_{SO_4}-\theta_D)$, where θ_D is the fraction of anion sites occupied by dye ions. It is practically independent of SO_4, since θ_{SO_4}, θ_D and θ_H do not vary very much. To a rough degree of approximation, therefore,

$$\psi = \text{constant}-\frac{RT}{2F}\ln SO_4 \tag{3}$$

for low concentrations of salt, and the potential is controlled by the amount of sodium sulphate added to the external solution. As the salt concentration is increased, however, θ_{Na} becomes more important. This leads to an increase in θ_{SO_4} which counteracts the external increase in sulphate-ion concentration, and results in a progressively slower rate of decrease of potential. Ultimately a limiting potential would be reached asymptotically if the terms $\Delta\mu_i$ were independent of the salt concentration. In practice, salting out of the non-polar part of a dye must occur and $\Delta\mu_i$ decrease, but for present purposes it is sufficient to realize that the fall in potential is only given by equation (3) at low salt concentrations (cf. figure (2) (I)).

The change in potential is mirrored by the change in the distribution of the dye ions in accordance with equation (1). Substituting from equation (2) into equation (1),

$$RT\ln\frac{\theta_i}{C_i} = \tfrac{1}{2}Z\,RT\ln\frac{\theta_{SO_4}}{SO_4}-(\Delta\mu_i-\tfrac{1}{2}Z\Delta\mu_{SO_4})+\tfrac{1}{2}(2-Z)\,RT\ln(1-\sum_i \theta_i). \tag{4}$$

If there are A sites per gram of keratin to be shared among sulphate and dye ions, each ion occupying only one site,

$$\theta_i = \frac{n_i}{A}$$

and
$$\theta_{SO_4} = \frac{n_{SO_4}}{A},$$

where n_i = number of dye ions adsorbed per gram of keratin, and n_{SO_4} = number of sulphate ions adsorbed per gram of keratin. Then

$$\ln\frac{n_i}{C_i} = -\tfrac{1}{2}Z\ln\frac{SO_4}{n_{SO_4}} - \left(\frac{\Delta\mu_i - \tfrac{1}{2}Z\Delta\mu_{SO_4}}{RT}\right) + \tfrac{1}{2}(2-Z)\ln(A - n_i - n_{SO_4}). \qquad (5)$$

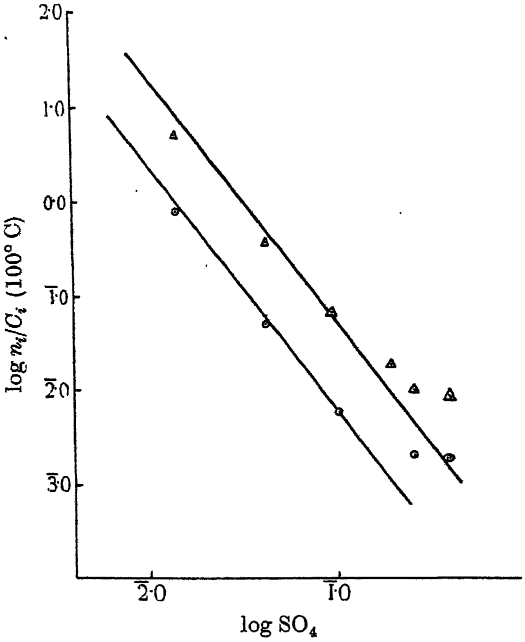


FIGURE 1. —— theoretical slope.

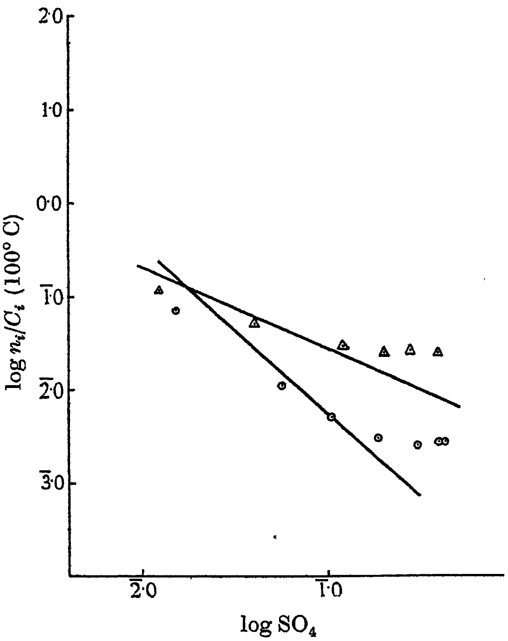
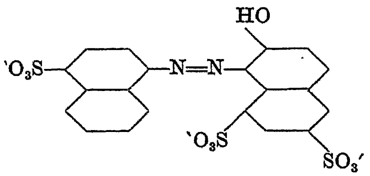
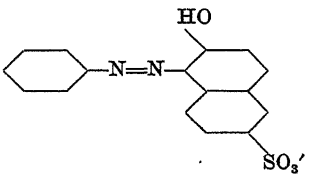


FIGURE 2. —— theoretical slope.

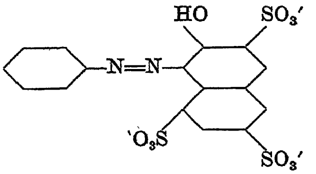
△ Dye 15



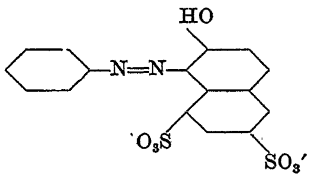
△ Dye 2



⊙ Dye 6



⊙ Dye 5



Under the conditions of Speakman & Clegg's experiments, n_{SO_4} and $(A - n_i - n_{SO_4})$ were fairly constant. Increasing quantities of sodium sulphate were added to the dye solution, kept at boiling-point, and measurement made of the amount of dye in the solution. From their results $\log(n_i/C_i)$ has been calculated, and this quantity plotted against $\log SO_4$ in figures 1–6, taking as an approximate measure of the latter the amount of sodium sulphate added per litre of solution.

It is evident that at low concentrations of salt equation (5) expresses the correct dependence of the slope of $\log(n_i/C_i)/\log SO_4$ on the valency of the dye ions (the theoretical slope is indicated in each case), but at higher concentrations departures occur in the expected direction.

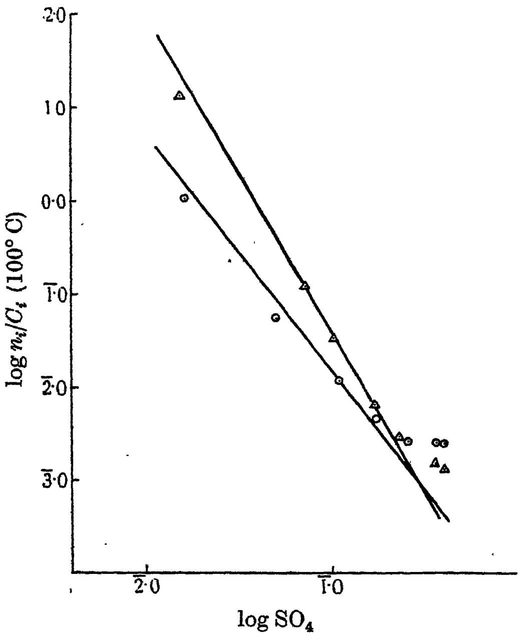
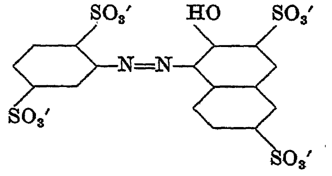
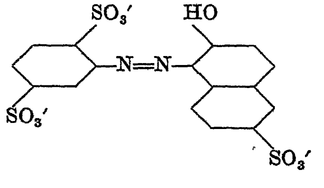


FIGURE 3. ——— theoretical slope.

△ Dye 11



⊙ Dye 10



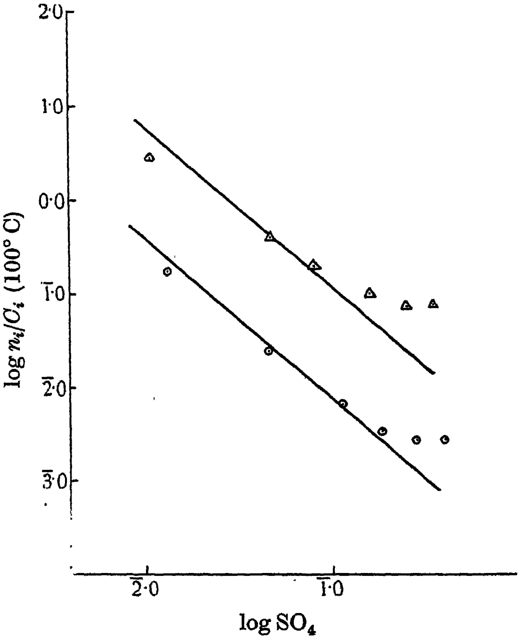
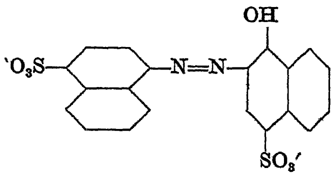
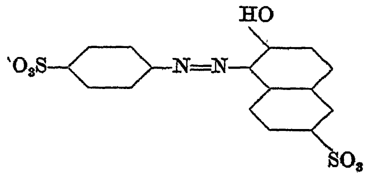


FIGURE 4. ——— theoretical slope.

△ Dye 39



⊙ Dye 8



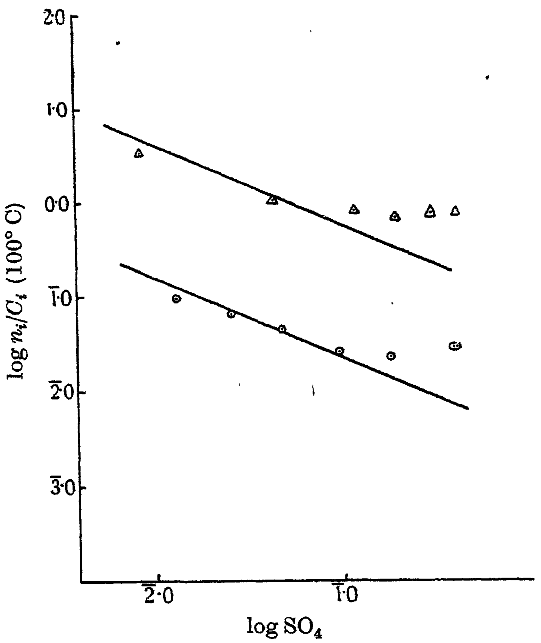


Figure 5. ——— theoretical slope.

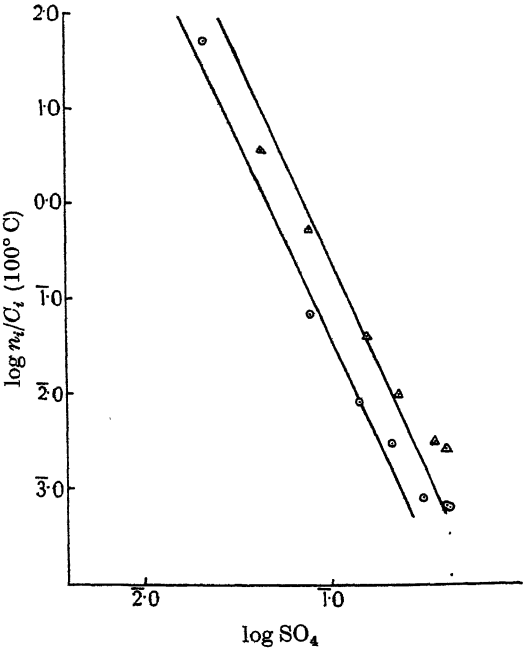
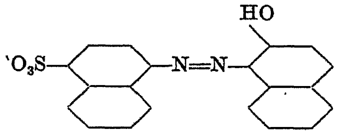
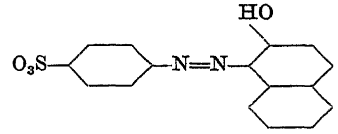


Figure 6. ——— theoretical slope.

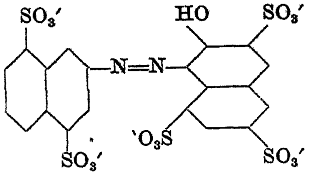
△ Dye 13



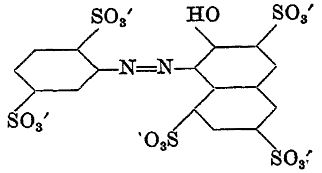
⊙ Dye 1



△ Dye 16



⊙ Dye 12



Influence of the constitution of the dye

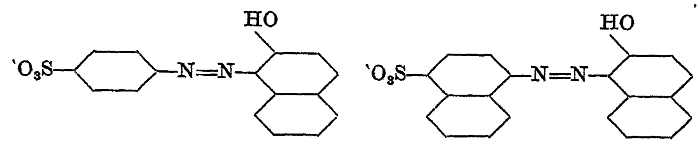
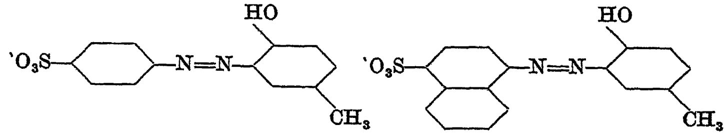
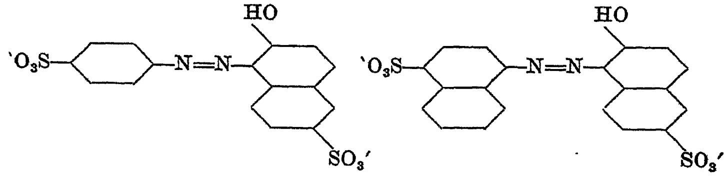
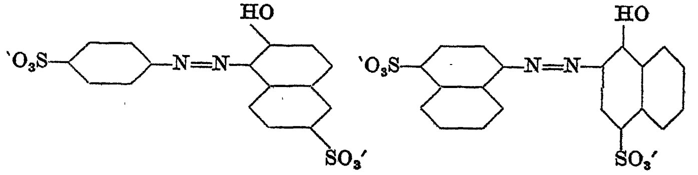
The curves allow the influence of a further factor to be seen. Dyes with the same valency Z are equally affected by changes in the non-specific force due to the fibre charge. 'Desorption' curves for such dyes are not necessarily superimposed, however, because each dye has an individual value for its intrinsic affinity $\Delta\mu_i$. In fact, the vertical displacement of one curve with respect to a parallel curve for another dye is a direct measure of the difference in $\Delta\mu_i$ for the two dyes. By relating such differences to the corresponding change in structure, it is possible to find what increment of chemical potential is conferred by particular elements of structure. Since the curves for a given valency are parallel, any pair of corresponding points can be selected to obtain this information. A number of dyes measured by Speakman & Clegg at only one concentration of sulphate ion can then be included. In table 1 the value of $RT \ln (n_i/C_i)$ has been calculated for a number of the simpler dyes in their paper. Their numbering of the dyes has been retained, and their paper should be consulted for details of the structures and purity of the dyes. Dye no. 3 has been omitted, since there is doubt about its identity.

TABLE 1

dye	$RT \ln (n_i/C_i)$ ($T = 100°$ C) kcal.	dye	$RT \ln (n_i/C_i)$ ($T = 100°$ C) kcal.
1	−1·52	17	−2·21
2	−1·59	18	−1·15
4	−2·62	19	−2·18
5	−2·55	20	−2·12
6	−2·71	22	−1·56
7	−2·57	23	−3·00
8	−2·55	24	−2·71
9	−2·57	25	−2·21
10	−2·62	26	−0·89
11	−2·90	27	−1·23
12	−3·17	28	−0·36
13	−0·19	31	−2·24
14	−1·05	32	−1·52
15	−2·07	39	−1·10
16	−2·56	40	−1·37

In table 2 pairs of dyes have been selected to obtain the increment corresponding to a benzene ring.

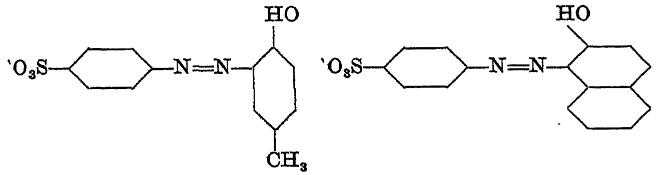
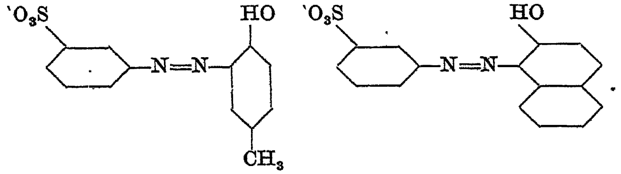
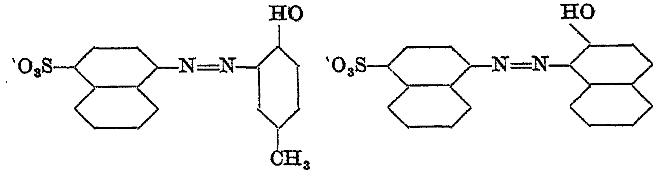
TABLE 2. INCREMENT FOR THE BENZENE RING

dyes compared	structures	increment in $\Delta\mu_i$ kcal.
1, 13		−1·33
17, 18		−1·06
8, 14		−1·51
8, 39		−1·46

Although the values of the increment are all of the same order, it cannot be said from these figures to what extent the increment depends on the molecule as a whole; and further work with refined methods of purification, and the elimination of the approximations from the terms in equation (5), would be needed for that purpose.

The substitution of a benzene ring for a methyl group gives a dye molecule of greater affinity, as can be seen from table 3.

TABLE 3

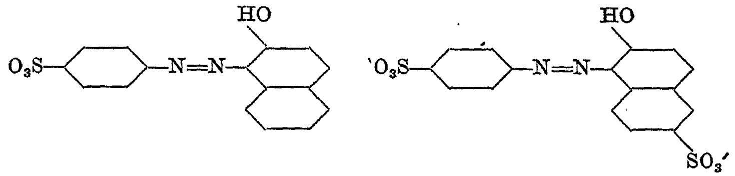
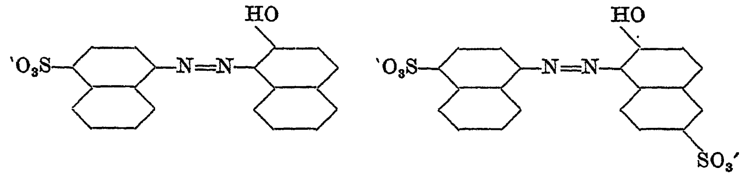
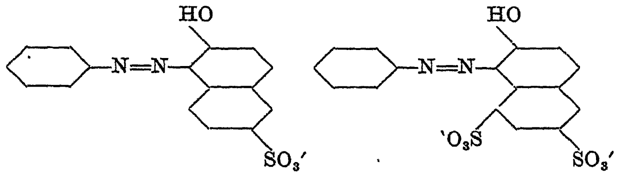
dyes compared	structures	increment in $\Delta\mu_i$ kcal.
17, 1		$-0{\cdot}69$
20, 22		$-0{\cdot}66$
18, 13		$-0{\cdot}96$

Comparison of the average increment here of $-0{\cdot}77$ kcal. with the previous average in table 3 of $-1{\cdot}34$ kcal. suggests that a methyl group corresponds to an increment of $-0{\cdot}57$ kcal.

A different situation arises when an —SO_3' group is added to a dye ion. The valency is increased by one, and the net effect on the adsorption of the dye ion is determined by the sum of the electrostatic contribution ψF, and the change in intrinsic affinity (i.e. in $\Delta\mu_i$). If the fibre potential is high (i.e. little sodium sulphate is present) the adsorption tendency of the dye is greatly increased because ψF is large and negative, but when much salt is present, the increase in the hydrophilic character of the dye, through the masking of a benzene ring, is no longer outweighed by the electrostatic term. The adsorption of the more highly charged dye ion is therefore less in strong salt solution than that of the less charged ion, in contrast to the position in dilute salt solution.

The decrease in $-\Delta\mu_i$ per SO_3' group seems to get less as the number of —SO_3' groups increases, but the evidence is too scanty to formulate rules as yet. The three comparisons given in table 4, however, made from measurements by Speakman & Clegg as before in solutions containing 130 g. of $Na_2SO_4 10H_2O$ per litre, show that the introduction of the second sulpho group reduces the intrinsic affinity (assuming ψ to be small) by almost the amount contributed by a ring.

TABLE 4

dyes compared	structures	increment ($\Delta\mu_i+\psi F$) kcal.
1, 8		1·03
13, 14		0·86
2, 5		0·96

II. THE TEMPERATURE COEFFICIENT OF THE IONIC EQUILIBRIA

The relations above hold for the particular temperature of the experiments. Frequently the value of $\Delta\mu_i$ for an ion is strongly temperature dependent, and the various equilibria are considerably altered between the temperature limits of 0 and 100° C (Wyman 1939; Steinhardt, Fugitt & Harris 1940). Some care is necessary as before in distinguishing between dissolved and fibrous proteins. In the adsorption of a single pure acid by a dissolved protein, it is known that if the anion has very little affinity for protein, hydrogen ions are adsorbed in the acid region of the titration curve practically without accompanying anions, and against an increasing positive potential. The differential heat of adsorption of the protons should therefore decrease with increasing saturation of the fibre. However, one would expect the magnitude of this effect to be small, from the very fact that the potential is not great enough to ensure the adsorption of anions.

Fibrous proteins, on the other hand, cannot adsorb protons without accompanying anions, and the sum of the heats of adsorption of protons and anions is obtained, instead of the heat of adsorption of the protons alone. The fibre potential does not affect this sum, as its contribution is equal and opposite for the cations and anions.

It should be noticed that this potential must vary with temperature, since it depends on the values of $\Delta\mu_i$ for the various ions.

In deriving heats of adsorption from the values of the chemical potential in the two phases, solution and fibre, it is simplest to consider first the adsorption of a single acid with an anion of type i and valency Z.

An intrinsic partial molar heat of adsorption ΔH_i can be defined for each ion which will correspond to the intrinsic affinity $\Delta\mu_i$, and which will be the part of the heat which does not arise from the fibre potential. The fact that the temperature coefficient of the chemical potential throughout the system must be one-valued can then be taken into account, if the chemical potential of the ion is to remain unique, and expressed for an anion by

$$\frac{\partial}{\partial(1/T)}\left\{\left(\frac{\mu_i^0}{T}\right)_{\text{fibre}}+R\ln\frac{\theta_i}{1-\theta_i}-\frac{ZF\psi}{T}\right\}=\frac{\partial}{\partial(1/T)}\left\{\left(\frac{\mu_i^0}{T}\right)_{\text{soln.}}+R\ln C_i\right\}. \tag{6}$$

Since, by the definition above,

$$\frac{\partial}{\partial(1/T)}\left(\frac{\mu_i^0}{T}\right)_{\text{fibre}}-\frac{\partial}{\partial(1/T)}\left(\frac{\mu_i^0}{T}\right)_{\text{soln.}}=\frac{\partial}{\partial(1/T)}\left(\frac{\Delta\mu_i}{T}\right)=\Delta H_i, \tag{7}$$

it follows that if θ_i is kept constant,

$$\Delta H_i-\frac{\partial(ZF\psi/T)}{\partial(1/T)}=R\frac{\partial\ln C_i}{\partial(1/T)}. \tag{8}$$

The analogous equation for the proton is

$$\Delta H_{\text{H}}+\frac{\partial(F\psi/T)}{\partial(1/T)}=R\frac{\partial\ln H}{\partial(1/T)}. \tag{9}$$

Eliminating $\frac{\partial(\psi/T)}{\partial(1/T)}$,

$$\Delta H_i+Z\Delta H_{\text{H}}=\Delta H_{\text{acid}}=R\frac{\partial}{\partial(1/T)}\{\ln C_i+Z\ln H\}, \tag{10}$$

where ΔH_{acid} stands for the heat of adsorption of a gram-molecule of acid. When $C_i=ZH$, as in solutions of pure acid,

$$\Delta H_{\text{acid}}=(Z+1)\,R\frac{\partial}{\partial(1/T)}(\ln H). \tag{11}$$

On the other hand, if C_i is kept constant by the addition of the sodium salt of the acid, and if sodium ions are not appreciably adsorbed,

$$\Delta H_{\text{acid}}=ZR\frac{\partial}{\partial(1/T)}(\ln H). \tag{12}$$

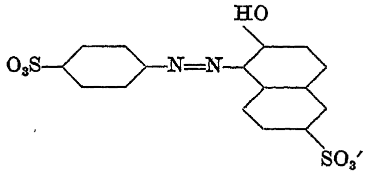
It can therefore be predicted that for univalent anions at a low but constant ionic strength, the change of pH with temperature for a given fibre content of acid is twice as great as it would be for solutions of pure acid, because the whole brunt of correcting for changes in $\Delta\mu_i$ and $\Delta\mu_{\text{H}}$ then falls on the hydrogen ions. Direct experimental evidence for this is lacking because the temperature coefficient is very small for the 'acid' titration curve, but it follows automatically if titration curves such as those for HCl, KCl mixtures in figure ((1), I) hold at different temperatures. Returning to the example presented by dyeing, it is seen that since the important ratio n_i/C_i is determined by the difference term $(\Delta\mu_i - \frac{1}{2}Z\Delta\mu_{SO_4})$, the temperature coefficient of n_i/C_i is determined largely by the corresponding term $(\Delta H_i - \frac{1}{2}Z\Delta H_{SO_4})$ or equivalently by $(\Delta H_{H_Z i} - \frac{1}{2}Z\Delta H_{H_2SO_4})$, where $\Delta H_{H_Z i}$ is the partial molar heat of adsorption of the dye acid $H_Z i$, and $\Delta H_{H_2SO_4}$ that of sulphuric acid.

When the experimental conditions are adjusted so that $(A - n_i - n_{SO_4})$ is constant, differentiation of equation (5) gives

$$\frac{\partial R}{\partial(1/T)} \ln\left\{\frac{n_i}{C_i}\left(\frac{SO_4}{n_{SO_4}}\right)^{\frac{1}{2}Z}\right\} = \frac{\partial}{\partial(1/T)} \frac{(\Delta\mu_i - \frac{1}{2}Z\Delta\mu SO_4)}{T} = \Delta H_i - \frac{1}{2}Z\Delta H_{SO_4}. \tag{13}$$

With properly controlled conditions, it is therefore possible to obtain the heat of replacement of a dye ion by an equivalent number of sulphate ions.

Equation (13) was tested experimentally for a relatively simple case. A pure sample of the non-aggregating dye below was provided by the research department of Messrs I.C.I. (Dyestuffs Group) Ltd., and used to dye a specimen of fine quality wool very kindly given by Professor J. B. Speakman.



In such experiments the temperature coefficient of the term $\frac{n_i}{C_i}\left[\frac{SO_4}{n_{SO_4}}\right]^{\frac{1}{2}Z}$ is required over as wide a range as possible. The problem of getting equilibrium at high temperatures must therefore be faced, in spite of decomposition of the keratin, and at low temperatures in spite of the slow diffusion of the dye ions. Some appreciation of the latter difficulty can be obtained from the study made by Speakman & Smith (1936) of the rate of adsorption by various types of keratin of Orange 2G (an isomer of the dye used here). These authors determined an apparent relative value of the diffusion coefficient k of this dye in terms of the initial value of $Q/\sqrt{t}$, where Q is the amount of dye adsorbed in time t. By analogy with diffusion into a semi-infinite slab, this ratio is at first constant and proportional to $k^{\frac{1}{2}}$, provided the surface concentration of dye remains constant. The latter condition appeared to be moderately satisfied. The values of $Q/\sqrt{t}$ obtained by these authors have been plotted against the reciprocal of the absolute temperature in figure 7. It is apparent that the diffusion of dye ions through keratin is an activated process, and one, moreover, requiring a high energy of activation. In the case of human hair this energy is of the order of 28 kcal., and for mohair about 23 kcal. A many thousandfold decrease in the diffusion coefficient therefore occurs if the temperature is lowered from 100 to 0° C.

The apparent decrease in the energy of activation at high temperatures in the presence of salt is due to the increasing unsaturation of the surface. The number of dye ions rather than their activity appears to be the relevant factor, and when the surface is unsaturated this number becomes temperature dependent. In N/10 Na_2SO_4 solution, the rate of dyeing does not increase above 60° C, probably because the rate of decrease of the number of dye ions at the surface (determined by a heat of exchange with sulphate ions of about 10 kcal.) just balances the increase in the diffusion coefficient ($Q/\sqrt{t}$ varying in accordance with 23/2 kcal.).

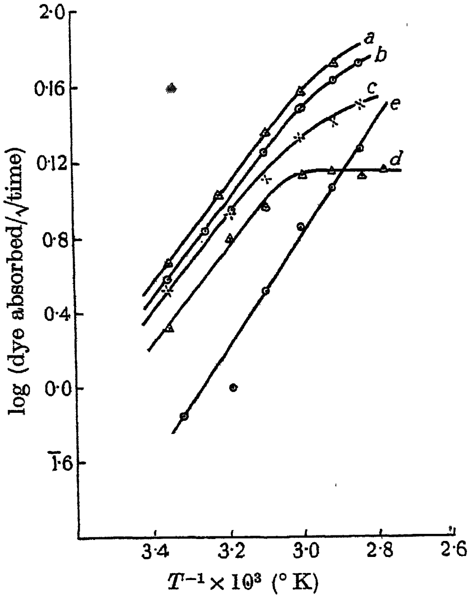
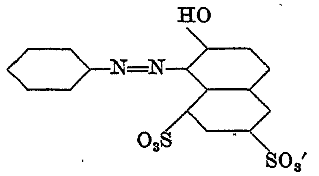


FIGURE 7. Data from Speakman & Smith (1936). (*a*) mohair: pH 1·11; (*b*) mohair: pH 2·04; (*c*) mohair: pH 2·14, N/50 Na_2SO_4; (*d*) mohair: pH 2·39, N/10 Na_2SO_4; (*e*) human hair: pH 2·04.

Activated diffusion of



The value of $\frac{n_i}{c_i}\left[\frac{SO_4}{n_{SO_4}}\right]^{\frac{1}{2}Z}$ varies with the origin and history of the wool fibre. It is therefore imperative to use the same sample of dyed wool for measurements of the temperature coefficient. A stock sample containing a known amount of dye acid and sulphuric acid can be prepared by dyeing at a fairly high temperature (e.g. 90° C) in the presence of salt, filtering at this temperature, centrifuging, and washing at 0° C with small quantities of water to remove decomposition products and salt. After centrifuging, brief warming restores complete equilibrium. Washing a part of this sample at 0° C with a small quantity of salt solution of a chosen strength, and centrifuging, then prepares a specimen for work in the given salt solution.

If now the temperature of such a treated sample is changed, the very small amount of residual solution immediately comes into new equilibrium owing to the minute amounts of dye and acid involved. This new equilibrium can be found experimentally by putting the fibre into a minimal amount of a solution previously adjusted to approximately the required concentration of dye, salt and acid. Final equilibrium is reached almost immediately, even at 0° C. The approximately adjusted solution is conveniently prepared by adding samples of the same dyed fibre at the given temperature to a solution containing the correct concentration of salt and then replacing them by the final sample. Indeed, one often finds, if the volume of solution relative to the fibre is kept small, that a single sample added to a salt solution of approximately the correct pH, will give within the time required for the temperature to become uniform, a concentration of dye that is unchanged by the addition of further samples. An alternative method, but one often not as convenient, is to flow salt solution through a column of dyed wool, when the upper part of the column plays the part of the preliminary samples.

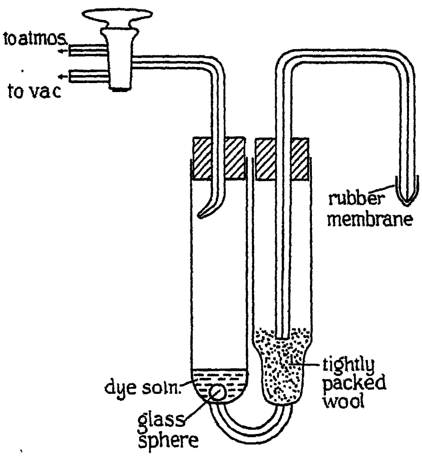


FIGURE 8. Micro-adsorption cell.

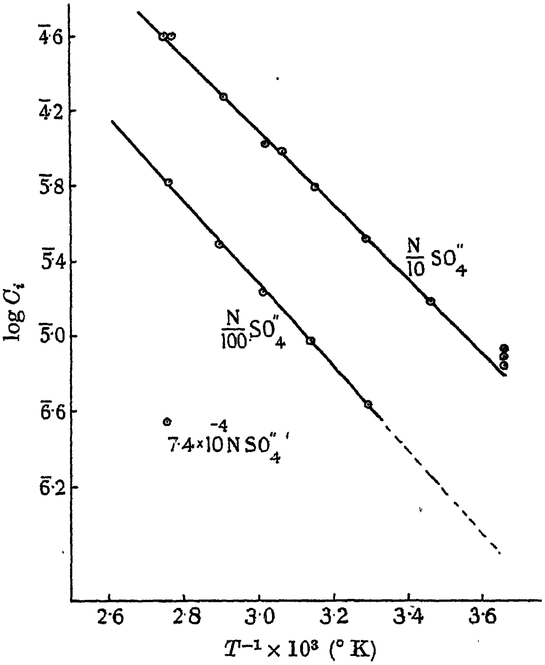
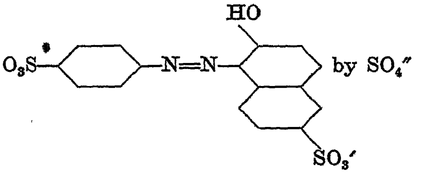


FIGURE 9. Heat of exchange.



An adsorption cell which was found suitable for equilibrating solution and fibre is drawn in figure 8. It was immersed in a thermostat, and arranged so that about 3 ml. of solution could be drawn back and forth through a tightly packed mat of fibre (*ca.* 0·5 g.). Removal of solution for analysis was carried out by increasing the pressure on the left-hand side (see figure) after removing the membrane over the fine capillary which leads from the arm containing the wool. The dye content of the solution was estimated polarographically by comparing the reduction wave height with that of standard solutions of dye.

Some measurements obtained by this null method, which avoids decomposition at high temperatures and lack of equilibrium at low, besides keeping the fibre content constant, are given in table 5. These results are presented graphically in figure 9. They are only exploratory, and do not possess the accuracy that could be attained by the methods outlined. However, the interesting result emerges that there is a large entropy of exchange. In terms of the partial molar heats and entropies of adsorption,

Table 5

H^+ per gram of fibre (dry weight) = $4{\cdot}05 \times 10^{-4}$ g.equiv.
SO_4'' per gram of fibre (dry weight) = $2{\cdot}79 \times 10^{-4}$ g.equiv.
Dye per gram of fibre (dry weight) = $1{\cdot}26 \times 10^{-4}$ g.equiv.

temp. T° C	$1/T$° K	dye concentration $c_i \times 10^{-6}$ M	log c_i	sulphate concentration SO_4 M
90	2·753	3·5	$\bar{6}$·545	$3{\cdot}7 \times 10^{-4}$
30·3	3·295	4·36	$\bar{6}$·640	$0{\cdot}5 \times 10^{-2}$
45·2	3·141	9·50	$\bar{6}$·978	
58·8	3·012	17·2	$\bar{5}$·236	
71·7	2·901	31·2	$\bar{5}$·494	
88·5	2·765	66·1	$\bar{5}$·820	
0	3·660	8·7	$\bar{6}$·940	$0{\cdot}5 \times 10^{-1}$
15·6	3·462	15·5	$\bar{5}$·190	
30·6	3·291	33·0	$\bar{5}$·519	
44	3·153	63·2	$\bar{5}$·801	
58	3·019	107	$\bar{4}$·029	
70·5	2·909	190	$\bar{4}$·279	
90·4	2·750	404	$\bar{4}$·606	
0	3·660	7·1	$\bar{6}$·851	$0{\cdot}5 \times 10^{-1}$
0	3·660	7·85	$\bar{6}$·895	
53·1	3·064	97	$\bar{5}$·987	
87·2	2·775	401	$\bar{4}$·603	

$$(\Delta\mu_i - \Delta\mu_{SO_4}) = (\Delta H_i - \Delta H_{SO_4}) - T(\Delta S_i - \Delta S_{SO_4}).$$

From equation (5), since $Z = 2$,

$$(\Delta\mu_i - \Delta\mu_{SO_4}) = -RT \ln \frac{n_i SO_4}{c_i n_{SO_4}},$$

from which it is found that at 90° C $(\Delta\mu_i - \Delta\mu_{SO_4})$ is approximately -3000 cal., whereas the above results show $(\Delta H_i - \Delta H_{SO_4})$ to be approximately $-10{,}000$ cal.

Hence $(\Delta S_i - \Delta S_{SO_4}) = -\frac{7000}{363} = -19$ cal./degree. Much greater restriction seems therefore to apply to the large heteropolar dye molecule within the fibre than to the smaller sulphate ions.

The author gratefully acknowledges continued and valuable discussion with Professor E. K. Rideal, and members of his Laboratory. He is indebted to Messrs I.C.I. (Dyestuffs) Ltd. for a grant which enabled this investigation to be made.

REFERENCES

Bancroft, W. D. 1914 *J. Phys. Chem.* **18**, 1, 118, 385.
Bancroft, W. D. 1915 *J. Phys. Chem.* **19**, 50, 145.
Gilbert, G. A. & Rideal, E. K. 1944 *Proc. Roy. Soc.* A, **182**, 335.
Hallitt, A. W. 1899 *J. Soc. Dy. Col., Bradford*, **15**, 30.
Speakman, J. B. & Clegg, G. 1934 *J. Soc. Dy. Col., Bradford*, **50**, 348.
Speakman, J. B. & Smith, S. G. 1936 *J. Soc. Dy. Col., Bradford*, **52**, 121.
Steinhardt, J., Fugitt, C. H. & Harris, M. 1940 *Bur. Stand. J. Res., Wash.*, **25**, 519.
Wyman, J. Jr. 1939 *J. Biol. Chem.* **127**, 1.

The transition effect of penetrating showers

By L. Jánossy and G. D. Rochester

The Physical Laboratories, The University, Manchester

(*Communicated by P. M. S. Blackett, F.R.S.—Received* 2 *March* 1944)

Observations of the transition effect of penetrating showers are reported and the effect is found to agree with the theory of Hamilton, Heitler & Peng as extended by Jánossy. The Z-dependence of the transition effect reported previously by Jánossy has not been confirmed.

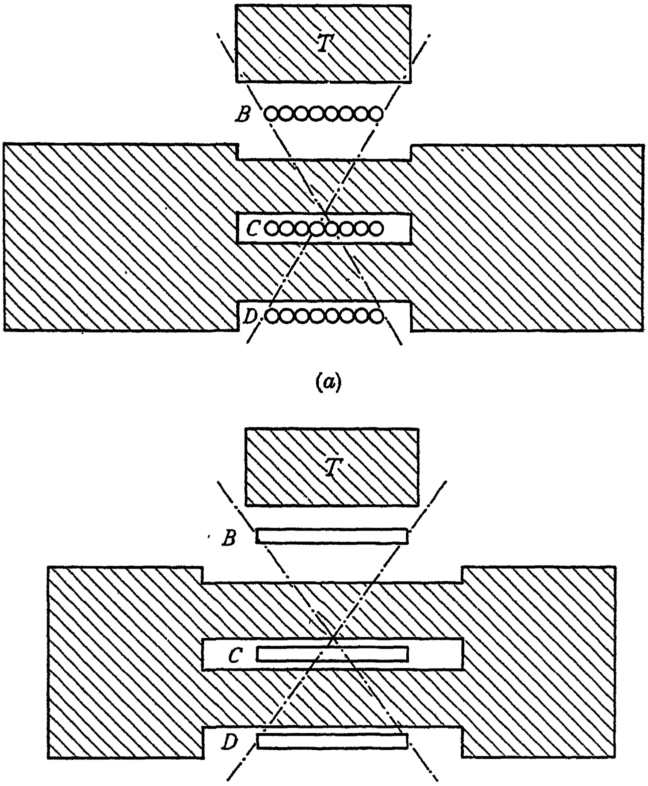
INTRODUCTION

For the interpretation of penetrating showers it is important to know the details of the transition effect accurately. We have therefore made measurements on the transition effect with a view to checking and improving upon previous results. The effect for lead was measured in some detail, but because of shortage of materials only one point was obtained for aluminium.

THE EXPERIMENTAL ARRANGEMENT AND THE RESULTS

The experimental arrangement is illustrated diagrammatically in figures 1*a*, *b*. It consists of three counter trays separated by layers of lead 15 cm. in thickness. As the arrangement is very similar to those used by Jánossy (1942) and by Jánossy

& Rochester (1943), a detailed description will not be given. The main difference from the arrangement used by Jánossy & Rochester is that the top tray of counters in the present arrangement contains eight counters instead of fifteen. The rate of triple coincidences $B_1 B_2 B_3$ is much reduced with eight counters, and therefore the rates of casual coincidences and triple knock-on showers are also much reduced without a corresponding decrease in the number of penetrating showers. It was found that the total rate of casual coincidences plus triple knock-on showers amounted to less than 5 % of the sevenfold coincidence rate.



(a)

(b)

Figure 1

Absorbers 46 × 46 cm.^{2} in area were brought close above tray B, and the rate of sevenfold coincidences as a function of the thickness of the absorber was determined. The area of the absorber was the same as in the previous arrangements.

Readings for all absorbers, taken over a period extending from December 1942 to November 1943, are given in table 1. They have all been reduced to a standard barometric pressure of 76 cm. by using the barometer coefficient (−11·7 % per cm. Hg), obtained by Jánossy & Rochester (1944). The results are plotted in figure 2.

The results given in the table or in any of the diagrams have not been corrected for the rate of accidental coincidences or knock-on showers, since in studying the transition effect we are mainly interested in the change of the shower rate and not in the actual rate. The variations in the rate of accidental coincidences plus knock-on showers is estimated to be less than one-third of the actual accidental rate, i.e. less than 2%.

TABLE 1

material	T		counts	time	rate
	cm.	g./cm.^{2}		hr.	counts per hour
light roof	0·00	0·00	216	1133·8	0·187 ± 0·013
lead	0·86	9·72	68	308·5	0·217 ± 0·026
	1·79	20·25	309	909·5	0·340 ± 0·019
	5·00	56·50	285	768·2	0·361 ± 0·021
	10·00	113·00	339	763·6	0·443 ± 0·024
	20·00	226·00	320	698·5	0·477 ± 0·026
aluminium	7·50	20·25	208	626·0	0·315 ± 0·022

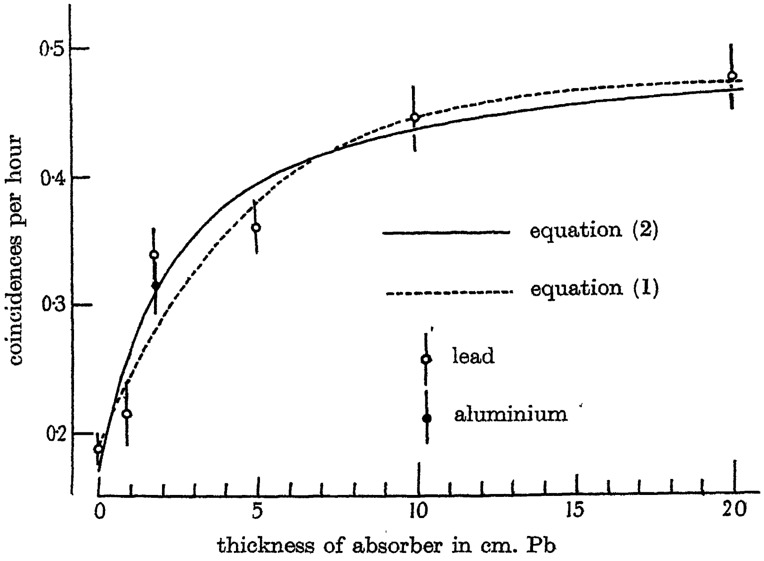


FIGURE 2. Transition curve observed (J. & R.) and calculated.

DISCUSSION

1. All points on the lead transition curve except the one for 1·8 cm. Pb can be fitted to an exponential function of the form

$$R = 0{\cdot}477 - 0{\cdot}29 \exp(-x/4{\cdot}6), \qquad (1)$$

where x is the thickness of the absorber in cm. of lead. Thus, disregarding the point $x = 1{\cdot}8$ cm. the transition effect can be accounted for in terms of a radiation of about 5 cm. range. This is exactly what is expected from the theory of Hamilton, Heitler & Peng (1943) and the theory put forward by Jánossy (1943).

Equation (1) is plotted as a broken line in figure 2. It is seen that the deviation from the dotted line of the point corresponding to 1·8 cm. is equal to three times the standard error of this point, and therefore it is possible that the deviation is not of statistical origin. The points are represented somewhat better by equation (2) which is plotted as a full line in figure 2:

$$R = 0{\cdot}46 - 0{\cdot}20 \exp(-x/4{\cdot}6) - 0{\cdot}09 \exp(-x). \tag{2}$$

Two interpretations can be offered for the deviation from the exponential function represented by (1), provided the deviation is taken to be real: (*a*) One can assume that only part of the penetrating showers is produced by primaries of range 5 cm. lead, the rest being produced by radiation of range approximately 1 cm. of lead. This second radiation might well consist of photons. (*b*) One can assume that some of the penetrating showers are accompanied by cascade showers. Because of the ordinary transition effect of cascade showers in lead a penetrating shower will have a larger chance of discharging the top tray of counters when it is covered with 2 cm. lead than in the case of an uncovered tray. The second interpretation would account better for the observed deviation.

2. In figure 3 the experimental points as observed by Jánossy (1942) are plotted together with the curve given by (2). The general shape of the lead transition curves found by Jánossy and by us are very nearly the same even though the two counter arrangements are different. This shows that the same phenomenon is being observed with the two arrangements. The maximum change in rate between $T = 0$ and $T = 10$ cm. is for Jánossy's arrangement (1942) 0·28 c.p.hr., for the arrangement of Jánossy & Rochester (1943) 0·25 c.p.hr., and for the present arrangement 0·26 c.p.hr. These rates must be considered the same when allowance is made for normal fluctuations.

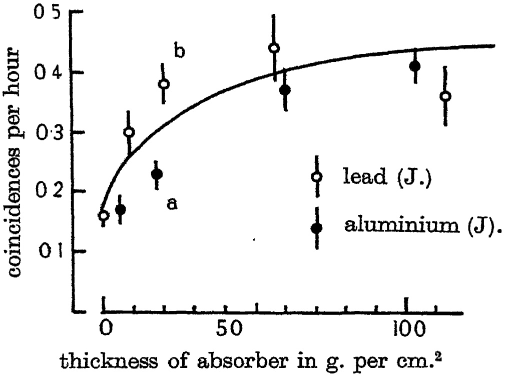


FIGURE 3

3. The one reading obtained for aluminium does not agree with Jánossy's value for this particular thickness of aluminium. We have made a careful study of this point for both lead and aluminium and have not been able to find a difference between the rates for the two materials. Throughout the whole of this experiment readings were staggered so as to eliminate slow changes in the apparatus, e.g. in the efficiencies of the counters, or seasonal changes in the radiation producing the showers. We conclude, therefore, that greater weight must be given to this result than to the former.

It is not known why the results of the two experiments differ, but it is unlikely that the difference can be due to different counter arrangements. Examination of figure 3 will show the previous deduction, that there is a difference between the rates for lead and aluminium, to be based chiefly on the points *a* and *b*. Because of statistical fluctuation the other points might well lie on the curve. The difference found by Jánossy might be due to an unusually large fluctuation in the point *a*.

REFERENCES

Hamilton, J., Heitler, W. & Peng, H. W. 1943 *Phys. Rev.* **64**, 78.
Jánossy, L. 1942 *Proc. Roy. Soc.* A, **179**, 361.
Jánossy, L. 1943 *Phys. Rev.* **64**, 345.
Jánossy, L. & Rochester, G. D. 1943 *Proc. Roy. Soc.* A, **182**, 180.
Jánossy, L. & Rochester, G. D. 1944 *Proc. Roy. Soc.* A, **183**, 186.

The barometer effect of penetrating showers

By L. Jánossy and G. D. Rochester

The Physical Laboratories, The University, Manchester

(*Communicated by P. M. S. Blackett, F.R.S.—Received 2 March* 1944)

The barometer coefficient of penetrating showers is evaluated from an extended series of observations and is found to be

$$B = -11\cdot7 \pm 2\cdot7\ \%\ \text{per cm. Hg.}$$

The significance of the correlation coefficient is discussed briefly.

1. Introduction

In the course of experiments on penetrating showers we had occasion to note the intensities of penetrating showers and the barometric pressures for a period of 16 months. Barometer correlation was not the main purpose of the experiments and therefore a number of different arrangements were used.

Though the method of correlating barometric pressure with the cosmic ray intensity is well known we sketch the procedure briefly in order to draw attention to certain points.

2. Calculation of the barometer coefficient

Denote by C_i the number of coincidences recorded in the ith reading. Denote similarly by t_i the duration of the ith reading and by b_i the average barometric pressure during this period. Introduce further the deviations of the individual values from the average according to the following formulae:

$$\gamma_i = C_i - t_i(\sum_k C_k)/\sum_k t_k, \quad \beta_i = b_i - (\sum_k b_k t_k)/\sum_k t_k. \tag{1}$$

The barometer coefficient is obtained as

$$B = \sum_k \gamma_k \beta_k / \sum t_k \beta_k^2, \tag{2}$$

or, if expressed in percentage change per cm. of Hg

$$\bar{B} = \frac{100}{76} \cdot \frac{B}{\Sigma C_k/n}, \tag{3}$$

where n is the number of readings.

3. THE CORRELATION COEFFICIENT

The question arises as to how far the barometer coefficient determined by (2) and (3) can be relied upon to express a real connexion between intensity and barometric pressure.

In this connexion the correlation coefficient, r, defined by equation (4) is often introduced,

$$r = \frac{\Sigma\beta_k\gamma_k}{\sqrt{(\Sigma\beta_k^2 t_k\,.\,\Sigma\gamma_k^2/t_k)}}. \tag{4}$$

For close correlation r has a value not very different from ± 1. It should be pointed out, however, that if we obtain a small value of r it does not necessarily mean that the value of the barometer coefficient is uncertain. Indeed, the value of r depends on the way in which the measurements are carried out.

Consider, for instance, a large number of short readings of the intensity. Because of statistical fluctuation individual readings will not follow closely the variation in the pressure. The correlation coefficient r will therefore be small.

It is clear that for a group consisting of a large number of readings individual fluctuations will not affect the average value, and therefore the barometer coefficient obtained from a large number of single readings will be reliable in spite of the large uncertainties in the single readings and the small value of r.

Similarly, one may consider variations which depend not only on the barometric pressure. In this case individual readings will not follow the barometer closely. Such fluctuations will, however, not influence the final result because they will cancel out on averaging over all readings, and the barometer coefficient obtained will have a real physical meaning. The correlation coefficient will be small in this case. The significance of this small value is that the intensity does not depend on the pressure only.

We see thus that the smallness of the value of r does not decide whether or not the corresponding correlation is significant.

4. STANDARD ERROR OF THE BAROMETER COEFFICIENT

The correct way to decide whether any experimental result is relevant or not is to calculate the standard error of the result from the error (or fluctuation) of the individual measurements which contribute to it.

The barometer coefficient can be regarded as a function of the measured quantities γ_i and β_i. We denote the standard errors of these quantities by $\delta\gamma_i$ and $\delta\beta_i$ and assume

$$\delta\gamma_1/\sqrt{t_1} = \delta\gamma_2/\sqrt{t_2} = \ldots \equiv \delta\gamma, \quad \delta\beta_1 = \delta\beta_2 = \ldots \equiv \delta\beta. \tag{5}$$

The standard errors of the average counting rate and the average pressure are of smaller order than those of the deviations γ_i and β_i and therefore will be neglected.

The standard error of the barometer coefficient can be obtained from (2) and (5) in the usual way. We find

$$\delta B/B = \surd\{\Sigma\gamma_k^2/t_k(\delta\beta)^2+\Sigma\beta_k^2 t_k(\delta\gamma^2)\}/\Sigma\gamma_k\beta_k. \tag{6}$$

If the error of the barometer readings is comparatively small we may put

$$\delta B/B = \delta\gamma\surd(\Sigma\beta_k^2 t_k)/\Sigma\gamma_k\beta_k. \tag{6a}$$

5. Amplitude of the non-barometric fluctuations

Besides establishing the correlation between barometric pressure and intensity it is also of interest to find out if the fluctuation of the corrected values is purely due to statistical fluctuation. If, for example, this fluctuation exceeds the statistical fluctuation one may assume that factors other than the barometric pressure influence the intensity.

We introduce the quantities

$$\eta_i = \gamma_i - B\beta_i t_i \quad (i = 1, 2, \ldots), \tag{7}$$

representing the deviations of individual readings from the average after having been reduced to average pressure. Regarding the η_i as functions of the γ_i one finds

$$\sum_k \eta_k^2/t_k = (\delta\gamma)^2(n-1), \tag{8}$$

and with help of equations (2), (5), (6a) and (8)

$$(\delta\gamma)^2 = (1-r^2)/(n-1). \tag{9}$$

The value of $(\delta\gamma)^2$ is the mean square deviation of the counting rates reduced to a standard pressure. The mean square deviation consists of two parts. One contribution is due to the statistical fluctuation and is given by $(\delta\gamma_{st})^2 = \Sigma C_k/\Sigma t_k$, and the second contribution is due to systematic changes of the intensity not due to barometric variation. We write for this second part $(\delta\gamma_{nst})^2 = \sigma^2$. The total fluctuation is thus

$$(\delta\gamma)^2 = \Sigma C_k/\Sigma t_k + \sigma^2, \tag{10}$$

and with help of (2), (7) and (9) we obtain from (10)

$$\sigma^2 = \frac{(\Sigma\gamma_k^2/t_k)(1-r^2)}{n-1} - \Sigma C_k/\Sigma t_k. \tag{11}$$

The above equation can be used to find the mean square amplitude σ^2 of the non-barometric variation. If σ^2 is small there is no indication of variations other than caused by changing pressure.

6. Numerical values

The values of the barometric coefficients, $\bar{B}$, obtained for the various arrangements are collected in table 1. The individual values have relatively large standard errors and we do not attach any importance to them.

The weighted mean of the fourteen individual results exceeds, however, four times its standard error, and therefore the existence of a barometer effect is ascertained.

The barometer coefficient thus obtained is much larger than the barometer coefficient observed for the total cosmic ray flux which is of the order of 3 % per cm. Hg.

The difference is probably even greater than appears from the figures, as approximately 30 % of the penetrating showers recorded by the arrangement '*I*' were knock-on showers. As the knock-on showers have the ordinary barometer effect the observed barometer effect of the mixture will be smaller than the effect of the penetrating showers alone. No correction for the admixture of knock-on showers has been made.

We conclude therefore that the barometer effect of penetrating showers must be assumed to be larger than that of the ordinary cosmic ray beam and is probably of the order of 10 % per cm. Hg.

No measurements of the barometer coefficient of penetrating showers have been carried out previously. Auger and Daudin (1942) and Cosyns (1940) have, however, determined the barometer coefficients of extensive air showers. The values found by these observers are between 10 and 20 % and are thus of the same order as the value found for the penetrating showers.

An attempt to interpret these effects is made in the following paper.

The value of σ^2 of the non-barometric variation was determined from equation (11) and averaged over the various arrangements. The value obtained is slightly

TABLE 1

		absorber cm. Pb					
number	reference	Σ	T	time hr.	count	$\bar{B}$	$(\delta\bar{B}/\bar{B})$
1	I	0	0	688	368	−5·3	1·1
2	I	0	3	102	92	−10·0	3·2
3	I	0	5	163	122	−15·1	1·0
4	I	0	10	381	311	−10·4	1·0
5	I	5	10	1290	1100	−12·9	0·43
6	I	35	10	1362	937	−4·3	1·4
7	I	0	3	206	145	+3·6	3·2
8	II	0	3	138	55	+15·9	2·0
9	II	0	0	965	198	−10·0	0·9
10	II	0	7·5*	626	208	−2·5	2·9
11	II	0	2	910	309	−18·6	0·42
12	II	0	1	309	68	−21·8	0·81
13	II	0	5	586	229	−10·5	0·80
14	II	0	10	495	230	−14·0	1·0

weighted average $\bar{B} = -(11{\cdot}7 \pm 2{\cdot}7)$ % per cm. Hg

I, Jánossy & Rochester (1943); II, Jánossy & Rochester (1944).

* cm. Al.

negative and therefore the fluctuation of the readings reduced to normal pressure is only statistical. There is no indication of a systematic variation due to causes other than variation of pressure.

Since the statistical fluctuation is large it cannot be excluded that systematic variations with amplitudes much smaller than the barometric fluctuation occur but are masked by the statistical fluctuation. Variations with an amplitude comparable to the fluctuations due to pressure changes would have been detected.

We are indebted to Dr T. G. Cowling of Manchester University and to Dr O. R. Frisch of Liverpool University for useful comment.

REFERENCES

Auger & Daudin 1942 *Phys. Rev.* **61**, 91.
Cosyns 1940 *Nature, Lond.*, **145**, 668.
Jánossy & Rochester 1943 *Proc. Roy. Soc.* A, **182**, 180.
Jánossy & Rochester 1944 *Proc. Roy. Soc.* A, **183**, 181.

Some aspects of the production of mesons and the barometer effect of penetrating showers

BY L. JÁNOSSY, *The Physical Laboratories, The University, Manchester*

(*Communicated by P. M. S. Blackett, F.R.S.—Received* 2 *March* 1944)

It is shown that the large barometer effect of penetrating showers can be understood in terms of an exponential absorption of the primaries giving rise to the showers. A number of other experimental results favour such an assumption. It is shown further that the exponential absorption of the primaries is compatible with the theory of Hamilton, Heitler & Peng when account is taken of fluctuations.

1. INTRODUCTION

From a large number of observations of penetrating showers Jánossy & Rochester (1944) have found the barometer coefficient of penetrating showers to be of the order of 10 % per cm. Hg. This barometer effect corresponds to an absorption of about 1 % per g. of the atmosphere. Jánossy & Rochester (1943*a*) have found, however, that the absorption of penetrating showers is only 0·1 % per g. of lead. Thus the absorption in air as deduced from the barometer effect appears to be ten times larger than that in lead.

As penetrating showers probably consist mainly of mesons, one would expect an absorption according to mass, and it would be very difficult to understand a much stronger absorption in air than in lead. The instability of the meson could hardly account for this behaviour in terms of decaying mesons, for the decay of mesons would give rise to a temperature rather than to a barometer effect.

The difficulty can be overcome by assuming that the barometer effect is not so much due to the absorption of penetrating showers in air but to the absorption of the primaries producing them. The large barometer effect and the small absorption of penetrating showers in lead can be understood in the following way.

Consider primaries which have just sufficient energy to penetrate the atmosphere at the normal pressure P. A fraction δN of the primaries will be absorbed in T (figure 1) giving rise to penetrating showers and thus will be recorded. If an extra absorber T_1 is placed on top of T, the fraction δN will give rise to penetrating showers in T_1 and not in T, but as the showers are only slightly absorbed in the absorber T they are almost as likely to be recorded when produced in T_1 as when produced in T; thus the absorber T_1 will have little effect on the observed intensity.

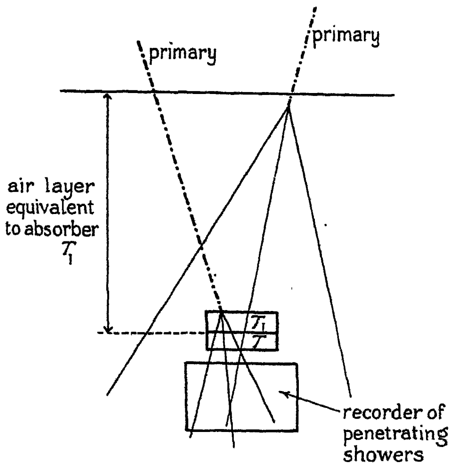

primary
primary
air layer equivalent to absorber T_1
T_1
T
recorder of penetrating showers

FIGURE 1. Showers produced in equivalent layers of lead and air.

Compare the effect of the extra absorber T_1 with the effect of an equivalent increase P_1 of the barometric pressure. P_1 should be so chosen that the extra air above the arrangement due to the pressure P_1 should have the same mass per unit area as the absorber T_1.

The effect of the increase of air pressure on the coincidence rate is very different from the effect of the absorber T_1. The fraction δN which could just reach the top of the apparatus when the air pressure was P will, when the pressure increases to $P+P_1$, be stopped in that layer above the surface at which, after the change, the pressure has now become P. Because of the low density of air this layer will have a height of the order of 100 m. for average changes of P. Showers produced at these heights will be very diffuse when they reach the apparatus and will therefore usually not be recorded (figure 1).

It follows that the dense absorber T_1 placed close above the apparatus produces a decrease according to the absorption of the *showers* in T_1, while an extra mass of air above the apparatus produces a decrease according to the absorption of the *primaries* in the air.

The large barometer coefficient seems therefore to indicate that the primaries giving rise to penetrating showers are strongly absorbed in air. The absorption coefficient of the primaries which is required to account for the observed barometer effect is of the order

$$1/\mu_P = 100\,\text{g. per cm.}^2. \tag{1}$$

2. ABSORPTION OF THE PRIMARIES (EXPERIMENTAL)

There is some experimental evidence to show that the primaries giving rise to penetrating showers are absorbed with an absorption coefficient μ_P of the order given in (1).

(*a*) The transition effect of penetrating showers shows saturation at approximately 5 cm. of lead. This saturation can be accounted for either by assuming that a large fraction of the incident primaries are absorbed in 5 cm. of lead or by assuming that the primaries are producing a number of penetrating secondaries in 5 cm. of lead sufficient to discharge the shower recording arrangement.

The second alternative was put forward by Hamilton, Heitler & Peng (1943) [quoted in the following as H.H.P.] and by myself (1943), but the barometer effect seems to favour the first interpretation. It will be shown in the following that the two points of view can be reconciled.

(*b*) It was shown by Jánossy & Rochester (1943) that part of the penetrating showers is produced by neutral particles. Further, Rossi & Regener (1940) and Jánossy & Rochester (1943*b*) have observed a penetrating non-ionizing radiation with a mean free path of about 100 g. per cm.^{2}. It is very probable that both of these neutral radiations consist of neutrons. The more energetic neutrons may be able to give rise to penetrating showers, while the bulk of the neutrons having too little energy for producing penetrating showers may give rise only to ionizing secondaries.

The intensity found by Rossi & Regener at Mt Evans (4500 m. above sea-level) was estimated by us to be 30–60 times larger than the intensity of neutral particles in Manchester. The ratio of the intensities suggests an absorption coefficient of the order of (1).

3. ABSORPTION OF PRIMARIES (THEORETICAL)

(*a*) Two main types of absorption phenomena have to be considered:

(i) *Catastrophic absorption.* The primary loses all its energy in its first collision. The absorption law in this case is essentially exponential and does not depend on the incident spectrum, provided the collision cross-section does not vary much with energy.

(ii) *Range absorption.* The primaries lose energy gradually. The more energetic primaries penetrate deeper and the absorption law is a function of the energy spectrum.

Intermediate processes may also occur.

Consider the case in which the primaries lose a constant amount β of energy per gram of absorber and assume β to be independent of energy. Assume an incident spectrum of the type

$$S(E) = (E_0/E)^\gamma \quad (E > E_0). \tag{2}$$

The incident spectrum of the primaries is likely to be of this type, since it is known that the meson spectrum is between wide limits of this type, and it is not likely that the spectrum of the primaries producing the mesons is very different from the secondary meson spectrum.

The energy spectrum at the depth x below the top of the atmosphere will then be

$$S(E,x) = E_0^\gamma/(E+\beta x)^\gamma \quad (E > E_0), \tag{3}$$

and, neglecting particles of energy less than E_0, the absorption function is given by

$$I(x) = E_0^\gamma/(E_0+\beta x)^\gamma. \tag{4a}$$

(*b*) The barometer coefficient near sea-level, i.e. at $x = H$, is given by

$$B = (H/I)\,(dI/dx)_{x=H}\ 1{\cdot}33\ \%\ \text{per cm. Hg}; \tag{5}$$

if $\beta H \gg E_0$ this reduces to

$$B = -1{\cdot}33\gamma\ \%\ \text{per cm. Hg (near sea-level)}. \tag{5a}$$

As γ is of the order of 2 the barometer coefficient is of the order of 3 % per cm. Hg. Such a value is observed for the bulk of cosmic rays.

Note that the coefficient B in (5*a*) is independent of the specific loss β. Therefore the barometer coefficient for the primaries producing the mesons should also be of the order of 3 % if they lose energy gradually.

(*c*) The barometric pressure at Mt Evans is 0·62 atm. Therefore the ratio of the intensities at Mt Evans and at sea-level is from (4*a*) with $\gamma = 2$, $E_0 \ll \beta H$

$$I_{\text{Mt Evans}}/I_{\text{sea-level}} = 0{\cdot}62^{-2} = 2{\cdot}6. \tag{6}$$

This value is far too small to be compatible with the estimated ratio from 30 to 60.

Thus, assuming continuous absorption of the primaries, one fails to account either for the ratio of neutron intensities on Mt Evans and at sea-level or for the large barometer coefficient found at sea-level.

Both effects would be accounted for immediately by assuming catastrophic loss. Replacing equation (4*a*) by

$$I(x) = \exp\left[-10x/H\right], \tag{4b}$$

the modified equations (5) and (6) give numerical values in good agreement with the experiments.

Using (4*b*) instead of (4*a*) it becomes also somewhat easier to account for the transition effect of penetrating showers.

It seems, therefore, that the experimental facts are better accounted for in terms of catastropic absorption than in terms of range absorption.

4. COMPARISON WITH THE THEORY OF HAMILTON, HEITLER AND PENG

In this theory it is assumed that the incident primaries are protons and possibly neutrons. It is shown that these fast heavy particles lose energy in inelastic collisions with protons and neutrons resulting in the emission of mesons. The rate of energy loss for primaries of energy E is given by

$$\beta_\mu = 0{\cdot}9 \log_e (0{\cdot}3E). \tag{7}$$

(All energies are given in units of μc^2, μ rest mass of the meson.) At first sight one should expect absorption according to range as a consequence of equation (7). In the following it will be shown, however, that owing to the large fluctuations of energy loss connected with the mechanism proposed by H.H.P. the absorption function is of the range type only for very large thicknesses, but for smaller thicknesses the absorption curve is more nearly represented by an exponential function.

Though the number of single encounters of a primary on its way through the atmosphere is large, these encounters are not distributed at random for, as I have shown, a primary particle passing through a single nucleus is likely to collide with several nuclear particles (Jánossy 1943). As a result of this crowding of the single collisions a primary is likely to lose large amounts of energy in a few collisions.

The energy required for a primary to penetrate the atmosphere is found with the help of equation (7) to be of the order of 7000, and it is estimated that this energy is lost in about 10 energetic collisions.

Primaries with energies below 7000/10 = 700 are therefore expected to be stopped in the first collision. A primary with energy much less than 700 is thus only likely to reach sea-level if it happens not to collide with any nucleus on its way. The probability of a primary not colliding once along H when its mean free path is $H/10$ is

$$e^{-10} \cong 1/20{,}000.$$

Thus one out of 20,000 low-energy primaries will reach sea-level. The fraction of the primaries having sufficient energy to penetrate the atmosphere without fluctuation is $(E_0/7000)^\gamma$, where E_0 is the smallest incident energy sufficient for producing penetrating showers at sea-level. Assuming $\gamma = 2$ and $E_0 = 50$ this fraction is also 1/20,000. Thus the number of low-energy primaries having reached sea-level by fluctuation is of the same order as the number of high-energy primaries reaching sea-level after having suffered average loss.

Roughly speaking the primaries behave as if they consisted of two independent components, one component being absorbed exponentially, the other having a range absorption curve. Down to a certain depth the exponential component is preponderant, while below this depth the fluctuations become unimportant and the absorption curve corresponds to the average loss.

Assuming that the exponential part of the absorption curve persists to a depth somewhat below sea-level, all experiments described in the last section can be accounted for in terms of the theory of H.H.P.

As the absorption with fluctuation is a rather complex process, it seems worth while to confirm these qualitative considerations by a more detailed calculation.

5. MATHEMATICAL TREATMENT OF THE FLUCTUATION OF ENERGY LOSS

Diffusion equation

In the following the total energy loss will be subdivided into two parts: (1) large losses: write for the probability of an energy loss in a single nuclear collision exceeding y along the path dx, $L(y)\,dx$; (2) small losses: it will be assumed that the loss along a path dx is equal to $\beta\,dx$, β being independent of the energy. The diffusion equation giving the change of the differential energy spectrum $D(E,x)$ can be written as

$$\frac{\partial D(E,x)}{\partial x} = -L(0)\,D(E,x)+\beta\frac{\partial D(E,x)}{\partial E}-\int_0^\infty D(E+y,x)\frac{dL(y)}{dy}\,dy. \tag{8}$$

Integrating by parts equation (8) can be transformed into

$$\frac{\partial D(E,x)}{\partial x} = \beta\frac{\partial D(E,x)}{\partial E}+\int_0^\infty \frac{\partial D(E+y,x)}{\partial E}\,L(y)\,dy; \tag{9}$$

a particular solution of this equation is

$$D_\epsilon(E,x) = \exp\left[-\epsilon E-\lambda(\epsilon)\,x\right] \tag{10}$$

with

$$\lambda(\epsilon) = \beta\epsilon+\epsilon\int_0^\infty e^{-y\epsilon}\,L(y)\,dy. \tag{11}$$

The general solution of equation (8) is therefore

$$D(E,x) = \int_0^\infty a(\epsilon)\exp\left[-\epsilon E-\lambda(\epsilon)\,x\right]d\epsilon, \tag{12}$$

where $a(\epsilon)$ has to be chosen so that the solution obtained satisfies a boundary condition, that is $a(\epsilon)$ has to satisfy equation (13),

$$D(E,0) = \int_0^\infty a(\epsilon)\exp\left(-\epsilon E\right)d\epsilon. \tag{13}$$

The general expression for $a(\epsilon)$ can be obtained from equation (13) by means of the Mellin transform. We are, however, only interested in the special case where the incident spectrum obeys a power law. In this case we have

$$D(E,0) = \gamma E_0^\gamma/E^{\gamma+1} \quad \text{and} \quad a(\epsilon) = (\epsilon E_0)^\gamma/(\gamma-1)!. \tag{14}$$

The solution of equation (8) satisfying the boundary condition (14) is therefore

$$D(E,x) = \frac{1}{(\gamma-1)!}\int_0^\infty (\epsilon E_0)^\gamma \exp[-\epsilon E - \lambda(\epsilon)x]\,d\epsilon. \tag{15}$$

Integrating equation (12) with respect to E from E_0 to infinity one obtains the integral spectrum. It is found that

$$S(E_0,x) = \frac{E_0}{\gamma!}\int_0^\infty (\epsilon E_0)^\gamma \exp[-\epsilon E_0 - \lambda(\epsilon)x]\,d\epsilon. \tag{16}$$

The values of $S(E_0,x)$ as obtained from equation (16) depend on the choice of the function $L(y)$. Before evaluating equation (16) for the case of energy loss due to meson production two extreme cases may be considered.

(*a*) $L(y) = \text{const.} = 1/\mu_P$. This represents a case in which the primaries are stopped in the first collision independently of their energy. It is found with the help of equation (11) that

$$\lambda(\epsilon) = \beta\epsilon + 1/\mu_P; \tag{11a}$$

introducing this expression into equation (16) one finds

$$S(E_0,x) = \frac{E_0^\gamma}{(E_0+\beta x)^\gamma}\exp[-\mu_P x]. \tag{16a}$$

Thus the absorption is essentially exponential. The barometer coefficient as defined in equation (5) will be found to be

$$B = -1{\cdot}33(\gamma + H\mu_P). \tag{5b}$$

It can be inferred from this equation that the observed barometer effect can be accounted for by assuming sufficiently large but infrequent losses for the primaries.

(*b*) $L(y)$ is very large for small values of y and is zero for larger values, so that the average rate of loss due to $L(y)$ is β_μ. In this case one finds

$$\lambda(\epsilon) = (\beta+\beta_\mu)\epsilon \tag{11b}$$

and

$$S(E_0,x) = \frac{E_0^\gamma}{(E_0+[\beta+\beta_\mu]x)^\gamma}; \tag{16b}$$

thus the absorption is essentially a range absorption; this case is incompatible with experiment.

(*c*) To treat the case of energy loss according to the theory of H.H.P. assume

$$L(y) = \begin{cases} 0{\cdot}9/(y-B) & (y \geqslant y_1), \\ 10\times 7{\cdot}10^{-3} & (y < y_1), \end{cases} \tag{17}$$

with $y_1 = 320$ and $B = 240$. A justification of this expression will be given in the Appendix. Further, assume $\beta = 0{\cdot}02$, this representing the ionization loss suffered by a proton.

Introducing equation (17) into equation (11)

$$\lambda(\epsilon) = \beta\epsilon - 0{\cdot}9\epsilon e^{-B\epsilon}\, Ei(-(y_1 - B)\epsilon), \tag{11c}$$

$Ei(-x)$ is the exponential integral. No analytic expression for equation (16) can be obtained in this case, but the integral can be evaluated by the saddle-point method.

The integrand in equation (14) vanishes both for small and large values of ϵ and has therefore maxima between.

Proceeding in the usual way, write for the integrand on the right-hand side of (16)

$$\text{integrand} = \exp[-f(\epsilon)], \tag{18}$$

$$f(\epsilon) = \epsilon E_0 + \lambda(\epsilon)\, x - \ln(\epsilon E_0); \tag{19}$$

saddle points are defined by

$$f'(\epsilon) = E_0 + \lambda'(\epsilon)\, x - \frac{1}{\epsilon} = 0 \quad \text{for} \quad \epsilon = \epsilon_0. \tag{20}$$

It is convenient to represent all quantities as functions of the independent parameter ϵ_0. Thus it is

$$x(\epsilon_0) = \frac{\dfrac{1}{\epsilon_0} - E_0}{\lambda'(\epsilon_0)}; \tag{21}$$

x is plotted as function of ϵ_0 in figure 2. This graph is rather remarkable. It is seen that in the region

$$320 < x < 1600, \tag{22}$$

$f(\epsilon)$ has three extrema, i.e. two minima and a maximum. The value S of the integral (16) for values of x in the interval given by equation (22) can be regarded as the sum $S_1 + S_2$ of the contributions of the two saddle points.

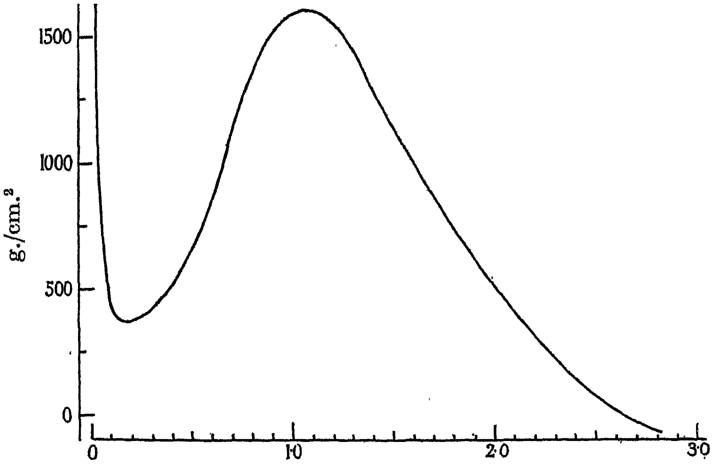


FIGURE 2. Position of saddle points.

Consider first the branch corresponding to the larger values of x and the small values of ϵ_0. From equation (11c) the following asymptotic expression valid for small values of ϵ can be derived

$$\lambda'(\epsilon) \approx \lambda(\epsilon)/\epsilon \approx \ln(140\epsilon). \tag{23}$$

Introducing this into equation (18) and carrying out the integration in the neighbourhood of the first saddle point, then the contribution of this saddle is

$$\left.\begin{aligned} S_1(E_0, x) &= 1/\pi\epsilon_0^2, \\ x &= 1/(-0{\cdot}9\epsilon_0 \ln(140\epsilon_0)). \end{aligned}\right\} \tag{24}$$

Comparing equations (24) and (7) it is noted that x is roughly equal to the range of the primaries with energy $1/\epsilon_0$, while S_1 is roughly the fraction of incident particles with energies exceeding $1/\epsilon_0$. Thus the branch of the curve (figure 2) corresponding to the small values of ϵ_0 represents the fictitious component showing range absorption.

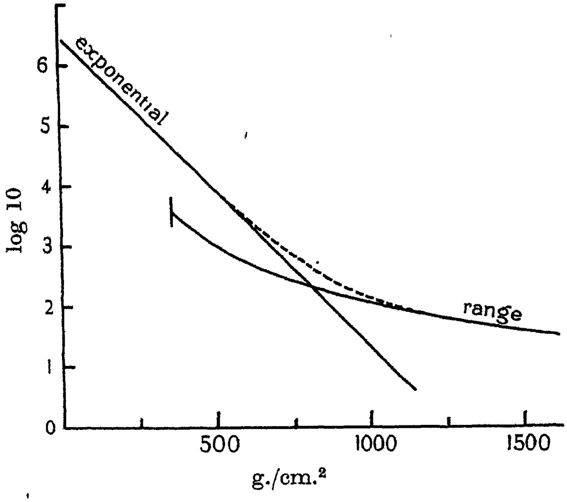


Figure 3. Absorption of primaries in the atmosphere.

The contribution S_2 due to the branch corresponding to the larger values of ϵ_0 has been evaluated numerically, and it is found to be very nearly exponential in x. Both contributions S_1 and S_2 as well as $S = S_1 + S_2$ are plotted in figure 3. It is seen from figure 3 that our expectations regarding the nature of the absorption curve in § 3 are confirmed by the quantitative treatment. The absorption is purely exponential down to a critical depth H_c, while for greater depth the curve is represented by a power function. The critical depth is, according to figure 3, about 800 g. per cm.^{2}. This value is somewhat short of the total depth of the atmosphere. Considering, however, the large uncertainties in the numerical values for the cross-sections involved one cannot attach any significance to the exact numerical value of this critical depth.

It can thus be concluded that the exponential absorption of the primaries should persist down to a critical depth which is of the same order as the height of the atmosphere.

6. COMPARISON WITH EXPERIMENT

(*a*) The ratio of the primary intensities on Mt Evans and near sea-level as calculated from figure 3 is about thirty, in agreement with the experimental estimate.

(*b*) Cloud-chamber photographs at high altitudes have been obtained by Powell (1940, 1941), Hughes (1941), Wollan (1941) and Hazen (1944). These photographs show groups of penetrating particles. No evidence is, however, found for the occurrence of more complicated processes such as one would expect due to energetic heavy particles capable of passing through more than one nucleus. These experimental findings are to be expected according to the present picture. The small groups of mesons can be attributed to the exponential component which must be expected to be preponderant at 3000 m. above sea-level. More complicated penetrating showers must be comparatively rare at these heights, as most of the energetic primaries are contained in the range component which is relatively weak there.

(*c*) The barometer coefficient near sea-level taken from the slope of the curve in figure 3 is found to be 7 % per cm. Hg.

This value is somewhat smaller than the observed value. The disagreement is, however, by no means serious, as the barometer coefficient changes very rapidly with depth in the region of the critical depth. Indeed, at heights above the critical depth where the absorption is nearly exponential the barometer coefficient as calculated from figure 3 is 16 % per cm. Hg. Along a comparatively short distance the value of the coefficient drops down to about 3 %. Thus a slight change in the estimation of H_c would result in a large change of the estimated value of the barometer coefficient at sea-level.

Barometer effect of extensive air showers

It can be shown easily that the exponential component consists mainly of low-energy particles, that is, of particles with energies of about 30–130. Extensive air showers observed by counters separated by a few metres have necessarily much higher energies and therefore the exponentially absorbed component cannot be responsible for the production of the extensive air showers.

The proposed theory gives therefore no explanation for the large barometer coefficients of extensive air showers observed by Auger & Daudin (1942) and Cosyns (1940). The large barometer coefficient of extensive air showers may be due to the absorption of the showers themselves and not to the absorption of the primaries producing them. A different view was put forward by Auger & Daudin (1942).

Energies of penetrating showers

To account for the barometer effect of penetrating showers near sea-level in terms of the exponentially absorbed component it is necessary to assume that about half of the showers at sea-level are due to the exponential component. This half of the showers should consist of showers with total energies not exceeding 130 and having often much lower energies.

Though at present no direct evidence supporting this prediction is available, it may be pointed out that this is quite compatible with my original estimation of the energies of penetrating showers based on observation (Jánossy 1942, page 369).

Latitude effect of penetrating showers

If the present interpretation is correct, penetrating showers should show a marked latitude effect. All the energies giving rise to the exponential component are forbidden at low latitude, and only the component due to high-energy primaries should be observed there. If the exponential component amounts to 50 % at high latitudes the latitude effect should also amount to 50 %.

Further, because of the absence of the exponential component the barometer effect at low latitudes should be of the order of 3 % only.

The shape of the transition curve for penetrating showers at low latitudes might also prove somewhat different from that at high latitudes.

The only penetrating shower experiments carried out at low latitudes are those of Wathaghin, Santos & Pompeia (1940). The coincidence rates observed by Wathaghin are of the same order as those recorded in Manchester, but as the arrangements used are rather different, no quantitative conclusions can be drawn.

Confirmation of these predictions would support the mechanism put forward in this paper. On the other hand, should these predictions prove wrong, one would be forced to assume that primaries of sufficient energy to penetrate the earth's magnetic field at low latitudes, i.e. primaries largely exceeding 600, still showed exponential absorption curve. Such behaviour might be accounted for by assuming that high-energy protons and neutrons suffer energy losses greatly exceeding those predicted by H.H.P. At present there is, however, no need for such an assumption.

7. APPENDIX

The differential cross-section for the emission of mesons in close collisions is given by H.H.P. as

$$d\Phi = (a/E^2)\,dE, \tag{25}$$

with $a = 1{\cdot}5 \times 10^{-24}$ cm.^{2}. This cross-section can be interpreted schematically by assuming that in a collision with the closest approach x a meson of energy

$$E = a/\pi x^2 \tag{26}$$

is emitted, provided x is not too large and provided E is less than the primary energy.

Two kinds of collisions have to be considered: (1) collisions where the fast particle crosses the actual area of the nucleus; (2) collisions where the fast particle by-passes the nucleus. In the second kind of collision the fast particle suffers distant collisions with all of the nuclear constituents, and it should give rise to the emission of slow mesons. In this case the assumption that the nuclear particles are independent is certainly not correct and the emission of mesons cannot be estimated in a simple way. It seems, however, reasonable to assume that in the case of the distant collision the nucleus acts more like a single body and the collision will be elastic with little loss of energy. We shall therefore neglect the loss of energy due to the distant collisions.

The energy loss in collisions where the primary passes through the nucleus is estimated roughly as follows. The radius r_A of a nucleus of weight A may be assumed as

$$r_A = 0{\cdot}53(A^{1/3}-1)(e^2/mc^2), \tag{27}$$

as shown elsewhere (Jánossy 1943). A distance r_0 can be so defined that a fast particle traversing a disk with radius r_0+r_A round the centre of the nucleus suffers in average only one collision having a distance of approach smaller than r_0. Ascribing an area πr_0^2 to each of the A nuclear particles one has

$$A\pi r_0^2 = \pi(r_0+r_A)^2, \tag{28}$$

and therefore $r_0 = r_A/3$ for $A = 16$.

Thus a fast particle traversing a nucleus will usually collide with one particle at a distance smaller than r_0 and suffer $A-1$ more distant collisions with the other particles. Averaging over all possible points of intersection of the fast particle with the nucleus I obtained with the help of equation (26) for $A = 16$ for this loss

$$y_1 = 240. \tag{29}$$

The energy y_2 which is lost in the closest collision has to be added. The probability per nucleus that the distance of this collision lies in the interval x, dx is $2\pi Ax\,dx$ with $x < r_0$. Therefore with the help of equation (26)

$$d\Phi_2(y_2) = Aa\,dy_2/y_2^2, \tag{30}$$

is obtained for the differential cross-section of a loss y_2. As the closest collision is assumed to have a distance less than r_0 it must be assumed

$$y_2 > a/\pi r_0^2 = 80. \tag{31}$$

With the help of equations (28), (30) and (31) one finds for the integral cross-section

$$\Phi_2(y_2) = \pi(r_0+r_A)^2\,80/y_2, \tag{32}$$

The total loss suffered is, however, $y = y_1+y_2$, and therefore the cross-section for the loss of energy exceeding y is

$$\Phi(y) = (r_0+r_A)^2\,80/(y-240) \quad (y \geqq 320). \tag{33}$$

Introducing numerical values into equation (33) one obtains finally,

$$L(y) = (N/A)\Phi(y) = 0{\cdot}9/(y-240) \quad (y \geqq 320). \tag{34}$$

Losses smaller than 320 occur according to our picture only when a particle of energy E less than 320 hits a nucleus. Such a particle is expected to lose the whole of its energy. Thus

$$L(y) = L(320) = 10{\cdot}7 \times 10^{-3} \quad (y < 320). \tag{35}$$

Equations (34) and (35) are valid for $y \leqq E$; they have to be replaced by $L(y) = 0$ for $y > E$.

The actual shape of $L(y)$ is not very important for the conclusions given here. It is, however, important for the conclusions that the losses due to the more distant collisions shall be neglected.

If one did not neglect the more distant collisions, but assumed that the energy loss in distant collisions was still given by the sum of the individual losses with the nuclear particles, one would obtain an additional term to $L(y)$. The effect of this additional term would be the same as if the ionization loss of the particles was increased about ten times. Thus with this extra term no particle with energy less than 300 could reach sea-level and therefore the whole exponential component would be cut out.

It seems, however, not unreasonable to cut off the small losses as has been done above.

I am indebted to Professor W. Heitler for commenting on this paper.

REFERENCES

Auger & Daudin 1942 *Phys. Rev.* **61**, 91.
Cosyns 1940 *Nature, Lond.*, **145**, 668.
Hamilton, Heitler & Peng 1943 *Phys. Rev.* **64**, 78.
Hazen 1944 *Phys. Rev.* **65**, 67.
Hughes 1941 *Phys. Rev.* **60**, 414.
Jánossy 1942 *Proc. Roy. Soc.* A, **179**, 361.
Jánossy 1943 *Phys. Rev.* **64**, 345.
Jánossy & Rochester 1943*a* *Proc. Roy. Soc.* A, **182**, 180.
Jánossy & Rochester 1943*b* *Proc. Roy. Soc.* A, **181**, 399.
Jánossy & Rochester 1944 *Proc. Roy. Soc.* A, **183**, 186.
Powell 1940 *Phys. Rev.* **58**, 474.
Powell 1941 *Phys. Rev.* **60**, 413.
Rossi & Regener 1940 *Phys. Rev.* **58**, 837.
Wathaghin, Santos & Pompeia 1940 *Phys. Rev.* **57**, 61.
Wollan 1941 *Phys. Rev.* **60**, 532.

Statistical thermodynamics of mixtures with zero energies of mixing

By E. A. GUGGENHEIM, M.A., Sc.D., *Imperial College of Science.*

(*Communicated by Sir Ralph Fowler, F.R.S.—Received* 3 *March* 1944)

A general formula has been obtained for the number of distinct arrangements of a mixture of any number of different types of molecules each with its own geometric properties. From this are deduced the thermodynamic properties of such mixtures when the energies of mixing are zero. In particular, the generalization of Raoult's law has been obtained. The technique used is considerably simpler than that previously applied to problems of this type.

1. *Introduction.* Van't Hoff's (1903) formulation of the laws of extremely dilute solutions was followed by a period of confusion, when these laws were inaccurately extrapolated to high concentrations. A new era began when G. N. Lewis (1908) defined a perfect solution as one which obeys Raoult's law in the form: the vapour pressure of the solvent is at constant temperature (and constant external pressure) directly proportional to the molecular fraction of the solvent. Lewis drew attention to the fact that mixtures exist, in particular benzene and ethylene chloride, which obey Raoult's law throughout the whole range of compositions, but he made no attempt to state the necessary conditions for this to occur. Shortly afterwards Washburn (1910) formulated the most important thermodynamic relations for a solution obeying Raoult's law throughout the whole range of compositions. He called such a solution 'ideal' and suggested that each real mixture should be compared with the ideal. He wrote: 'Wir können ziemlich sicher sein, dass eine Abweichung von der Forderungen dieser Theorie in einem gegebenen Fall auf physikalischen oder chemischen Ursachen beruht... und dass sie nicht das Resultat von Bemühungen darstellt, eine Gruppe von unvollkommenen Gesetzen anzuwenden, die für keine Art von koncentrierten Lösungen... gelten', and 'Damit die Gesetze der Idealen Lösung anwendbar seien, darf die Natur des Mediums oder das Kraftefeld in dem sich die Moleküle in der Lösung befinden, nicht sehr verschieden von demjenigen der reinen Flüssigkeit selbst sein.' Beyond this Washburn refrained from attempting a precise formulation of the necessary conditions for a mixture to be ideal.

It is easily shown thermodynamically that two or more substances can form a perfect or ideal mixture, only if they mix, at constant temperature and pressure, without any change of energy or volume. These conditions, known to be necessary, came to be regarded as sufficient (Hildebrand 1924) for ideality, until the contrary was pointed out (Guggenheim 1933) and the name 'semi-ideal' was proposed for mixtures having zero energy of mixing and zero volume change on mixing, but not obeying Raoult's law.

In 1932 I gave a statistical mechanical derivation (Guggenheim 1932) of the laws of ideal solutions requiring the assumption, among others, that the several kinds of molecule may be treated as spheres of at least approximately the same size, so that one could consider two configurations differing from each other only by the interchange of two molecules of different kinds. At a discussion held in Edinburgh on 'Liquids and Solutions' I pointed out (1937) that there were no grounds for believing that mixtures of molecules differing greatly in size would obey Raoult's law. Following this discussion Fowler & Rushbrooke (1937) applied statistical mechanics to a mixture of two kinds of molecules A and B, the former each occupying two sites of a lattice and the latter each occupying one. They obtained formulae valid for mixtures dilute with respect to A and mixtures dilute with respect to B, showing that Raoult's law is not obeyed even when the energy of mixing is zero. The essentially similar problem of adsorption of molecules each occupying two sites on a surface lattice was then taken up by Chang (1939), who obtained explicit formulae valid throughout the whole range of concentrations, by using the mathematical technique developed by Bethe for treating order-disorder transitions in alloys. Chang's results were adapted to mixtures of double and single molecules by Fowler & Guggenheim (1939), and the laws relating partial vapour pressure to composition for such mixtures were derived. Chang's method of approach was subsequently applied by Miller (1942, 1943) to mixtures of two kinds of molecules each occupying respectively three sites and one site. Miller also wrote down by analogy, but without proof, correct formulae for a mixture of chain-like molecules each occupying r sites and molecules each occupying a single site.

Chang's formulae were published just in time to be quoted and used by Fowler & Guggenheim (1939, p. 367), but owing to the war it is only recently that I have had occasion to study Chang's papers in detail. It has already been mentioned that Chang used a technique invented by Bethe for the study of order-disorder in alloys. Now this technique is superior to the incomparably simpler technique previously applied to these problems by Bragg and Williams only in that due account is taken of deviations from randomness due to the non-zero energy of mixing, an effect neglected by Bragg and Williams. It would therefore be surprising if the complicated technique of Bethe, invented to take account of a second-order effect of a non-zero energy of mixing, were really required in a problem in which the energy of mixing is assumed to be zero. I accordingly became suspicious that Chang's results should be derivable by simpler methods. This suspicion became a conviction, when I found that most of Chang's formulae could be transcribed into alternative forms exhibiting a simplicity and symmetry previously latent. Once convinced that a simple derivation existed, I had little difficulty in finding it.

The object of this paper is accordingly to give an elementary derivation of the thermodynamic formulae for systems with zero energies of mixing and zero volume changes of mixing. In the first place I shall consider mixtures of two kinds of molecules each occupying respectively r sites and one site, deriving the formulae already obtained by Miller by analogy, but without proof. Chang's formulae are immediately derivable as the simplest special case. According to the new derivation, Miller's formulae, proposed for simple chain molecules, are equally applicable to molecules with branched chains. In the second place I shall indicate briefly how the new simplified treatment leads immediately to formulae for mixtures of any number of kinds of molecules occupying any number of sites. The application of Bethe's method to systems of such complexity would be a formidable task and it is good to know that it is unnecessary.

2. *Statement of problem and notation.* Consider a system of N_A molecules A each occupying r sites and N_B molecules B each occupying one site, the total number of sites being $N_S = rN_A + N_B$. Denote by z the number of sites which are neighbours of any given site. Then the number of sites which are neighbours of a B molecule is of course z, but the number of sites which are neighbours of an A molecule is not rz (except in the trivial case $r = 1$), but say qz, where $q < r$ because some of the sites adjoining a site occupied by an element of a molecule A are occupied by the next element of the same molecule. Provided the A molecules are simple chains or branched chains, without any closed rings, the relation between q and r is

$$qz = rz - 2r + 2. \tag{2·1}$$

The exclusion of molecules having closed rings is an essential condition for the ensuing argument and the formulae obtained are therefore not applicable to molecules with closed rings. In the case of flexible open chain molecules, which can bend back on themselves, the ensuing argument is applicable provided q is still defined by (2·1), but the definition in words of q has to be modified as follows: qz is the number of sites neighbours of the r sites occupied by a molecule A, excluding those neighbours of each site occupied by the *next* element of the same molecule.

It is a fundamental assumption in the present problem that the energies of mixing are zero; in other words, that all configurations have the same configurational energy. The more complicated problem, when the energies of mixing are not zero, will be treated in the following paper (p. 213).

3. *Essential statistics.* Since all configurations are assumed to have the same energy, it follows that at a given temperature they all have the same statistical weight. This is equivalent to the statement that the arrangement of the molecules will be completely random. A number of consequences follow immediately from this randomness, in particular the following.

The frequency of occupation of a chosen site by a B molecule is $N_B/(rN_A + N_B)$, and the frequency of its occupation by an A molecule is $rN_A/(rN_A + N_B)$.

If a site is occupied by a B molecule, then the chances that a given neighbouring site is occupied by another B molecule or by an A molecule are as N_B is to qN_A.

If a site is occupied by an A molecule, then the chances that a given neighbouring site is occupied by another element of the same molecule or by another molecule (A or B) are as $r - q$ is to q.

If a site is occupied by an A molecule, then the chances that a given neighbouring site is occupied by a B molecule or by a different A molecule are as N_B is to qN_A.

From these follow immediately the following expressions for the frequencies of occupation of a chosen pair of sites in alternative ways:

Both by B $$\frac{N_B}{(rN_A+N_B)}\frac{N_B}{(qN_A+N_B)};\tag{3·1}$$

First by B, second by A $$\frac{N_B}{(rN_A+N_B)}\frac{qN_A}{(qN_A+N_B)};\tag{3·2}$$

First by A, second by B $$\frac{qN_A}{(rN_A+N_B)}\frac{N_B}{(qN_A+N_B)};\tag{3·3}$$

Both by different A's $$\frac{qN_A}{(rN_A+N_B)}\frac{qN_A}{(qN_A+N_B)};\tag{3·4}$$

Both by same A $$\frac{(r-q)\,N_A}{(rN_A+N_B)}.\tag{3·5}$$

The correctness of these values is readily verified by checking the values of the ratio of (3·1) to (3·2), of (3·3) to (3·4) and of (3·5) to the sum of (3·3) and (3·4); also by checking the values of the sum of (3·1) and (3·2) and of the sum of (3·3), (3·4) and (3·5). A thorough study of these expressions and the relations between them is profitable, although we shall in fact require to use only (3·1).

Now consider a particular set of r sites so interrelated that it is possible for them to be occupied by an A molecule. If the A molecules are flexible, the set of r sites must be such that an A molecule occupying them is not bent back on to itself. According to the assumption of randomness the frequency of occupation of this set of sites by an A molecule must be directly proportional to N_A, the total number of such molecules. This frequency can be denoted by KN_A/N_S, where K depends on z, r, q but is independent of N_A, N_B. The frequency of occupation of the whole set of sites considered by B molecules will also be required. By an obvious extension of (3·1), this frequency is

$$\frac{N_B}{(rN_A+N_B)}\left(\frac{N_B}{qN_A+N_B}\right)^{r-1}.\tag{3·6}$$

4. *Kinetic argument.* According to the principle of detailed balancing, when a system is in equilibrium any single process balances the inverse process. It is therefore possible to derive correct equilibrium conditions by considering an arbitrary elementary process and its inverse, even though the elementary processes considered may be so extremely infrequent, compared with some other elementary mechanisms, that they do not appreciably contribute to the attainment or maintenance of equilibria. Making use of this principle one can obtain an extremely simple kinetic derivation of the equilibrium conditions sought.

The elementary processes chosen for consideration are:

(*a*) the removal ('evaporation') of a molecule A from a selected set of r sites and its simultaneous replacement ('condensation') on these sites by r molecules B from the gas phase;

(*b*) the inverse process of removal ('evaporation') of r molecules B from the same set of sites and their simultaneous replacement ('condensation') by a molecule A from the gas phase.

The rate of process (*a*) will be directly proportional to the rth power of p_B, the partial pressure of B in the gas phase and to the frequency of occupation of the set of sites by an A molecule. According to the previous section, the rate will therefore be directly proportional to

$$p_B^r \frac{rN_A}{rN_A+N_B}. \tag{4·1}$$

The rate of process (*b*) will be directly proportional to p_A, the partial pressure of A in the gas phase, and to the frequency of occupation of the whole set of sites by r molecules B. According to (3·6), the rate will therefore be directly proportional to

$$p_A \frac{N_B}{(rN_A+N_B)} \left(\frac{N_B}{qN_A+N_B}\right)^{r-1}. \tag{4·2}$$

At equilibrium the rates of the elementary processes (*a*) and (*b*) must be equal and consequently

$$C' \frac{p_A}{p_B^r} = \frac{rN_A}{N_B} \left(\frac{qN_A+N_B}{N_B}\right)^{r-1}, \tag{4·3}$$

where C' depends on the characteristics of the molecules A and B; in particular on r, q and z; and finally on the temperature T; but C' is independent of N_A and N_B.

It will be convenient to make considerable use of thermodynamic functions denoted by λ, related to Gibbs' potentials μ by $\mu = kT \log \lambda$, and conveniently referred to by the name 'absolute activities' (Fowler & Guggenheim 1939, p. 66). The relation between the partial pressure p of a perfect gas and its absolute activity λ is

$$\lambda = p/kT\phi(T), \tag{4·4}$$

where $\phi(T)$ is the partition function of the gaseous molecule with the volume factor removed. All that matters here is that the ratio λ/p is independent of p. Therefore (4·3) may be replaced by the similar relation

$$C \frac{\lambda_A}{\lambda_B^r} = \frac{N_A}{N_B} \left(\frac{qN_A+N_B}{N_B}\right)^{r-1}, \tag{4·5}$$

where C, like C', depends on the characteristics of the molecules A and B and on the temperature, but is independent of N_A and N_B.

If f_A and f_B are used to denote the partition functions of molecules A and B respectively attached to specified sites, then $\lambda_A f_A/N_A$ and $\lambda_B f_B/N_B$ are configurational factors related to the geometry of the molecules and of the lattice. Since, moreover, only systems in which all lattice configurations have equal energy are being considered, $\lambda_A f_A$ and $\lambda_B f_B$ may be expected to be independent of temperature. It is accordingly convenient to replace (4·5) by the equivalent relation

$$\rho\frac{\lambda_A f_A}{(\lambda_B f_B)^r} = \frac{N_A}{N_B}\left(\frac{q_A N_A + N_B}{N_B}\right)^{r-1}, \tag{4·6}$$

where ρ, like C and C', is independent of N_A and N_B but may depend on r, q and z. It may, however, be expected, and will later be verified, that ρ, unlike C and C', is independent of the temperature.

The reader who is thoroughly conversant with the use of grand partition functions and the part played in these by the absolute activities λ, could probably write down (4·6) directly without need of the kinetic argument. He could perhaps also write down the exact form of the constant ρ, but he would probably have difficulty in giving a rigid proof. I therefore prefer to leave the form of ρ indefinite at this stage. All the required thermodynamic formulae can be derived from (4·6).

5. *Gibbs function and free energy.* According to the definitions of the absolute activities λ_A, λ_B and the partial potentials μ_A, μ_B these are related to the Gibbs function G at constant temperature and pressure according to

$$dG = \mu_A dN_A + \mu_B dN_B = kT\log\lambda_A dN_A + kT\log\lambda_B dN_B. \tag{5·1}$$

In order to make use of (4·6) in the integration of (5·1), one transforms the independent variables from N_A, N_B to N_S, N_A using

$$N_B = N_S - rN_A. \tag{5·2}$$

Differentiating (5·2) and substituting into (5·1), one obtains

$$dG = (\mu_A - r\mu_B)\,dN_A + \mu_B dN_S, \tag{5·3}$$

so that

$$\left(\frac{\partial G}{\partial N_A}\right)_{N_S} = \mu_A - r\mu_B = kT\log\frac{\lambda_A}{\lambda_B^r}. \tag{5·4}$$

Substituting from (5·2) into (4·6) and using the result in (5·4), one obtains

$$\frac{1}{kT}\left(\frac{\partial G}{\partial N_A}\right)_{N_S} = \log N_A - r\log(N_S - rN_A) + (r-1)\log(N_S - [r-q]\,N_A) - \log f_A + r\log f_B - \log\rho. \tag{5·5}$$

When (5·5) is integrated the integration constant is determined by the condition that when N_A is zero the formula for G must reduce to

$$\frac{G}{kT} = -N_S\log f_B \quad (N_A = 0). \tag{5·6}$$

Integrating (5·5) subject to the condition (5·6) one obtains for the Gibbs function G and the free energy F, assuming the pressure to be small,

$$\frac{F}{kT}=\frac{G}{kT}=\log N_A!+\log\frac{(N_S-rN_A)!}{N_S!}-\frac{r-1}{r-q}\log\frac{(N_S-[r-q]\,N_A)!}{N_S!}$$
$$-N_A\log\rho-N_A\log f_A-(N_S-rN_A)\log f_B. \tag{5·7}$$

Now transforming back to the variables N_A, N_B one obtains

$$\frac{F}{kT}=\frac{G}{kT}=\log N_A!+\log N_B!-\frac{r-1}{r-q}\log(qN_A+N_B)!+\frac{q-1}{r-q}\log(rN_A+N_B)!$$
$$-N_A\log\rho-N_A\log f_A-N_B\log f_B. \tag{5·8}$$

6. *Partial potentials and vapour pressures.* All the important thermodynamic properties of the mixture can be obtained from (5·8) by differentiation. If the subscript 0 is used to denote the value of a quantity for the substance A or B in its pure state, then one obtains for the partial potentials μ_A, μ_B, the absolute activities λ_A, λ_B and the partial vapour pressures p_A, p_B the formulae

$$\exp\left(\frac{\mu_A-\mu_{A,0}}{kT}\right)=\frac{\lambda_A}{\lambda_{A,0}}=\frac{p_A}{p_{A,0}}=N_A\frac{(N_A+N_B/r)^{\frac{r(q-1)}{r-q}}}{(N_A+N_B/q)^{\frac{q(r-1)}{r-q}}}, \tag{6·1}$$

$$\exp\left(\frac{\mu_B-\mu_{B,0}}{kT}\right)=\frac{\lambda_B}{\lambda_{B,0}}=\frac{p_B}{p_{B,0}}=N_B\frac{(rN_A+N_B)^{\frac{q-1}{r-q}}}{(qN_A+N_B)^{\frac{r-1}{r-q}}}. \tag{6·2}$$

It is readily verified that (6·1) and (6·2) satisfy (4·3) and the necessary relation

$$\left(\frac{\partial\mu_A}{\partial N_B}\right)_{N_A}=\left(\frac{\partial\mu_B}{\partial N_A}\right)_{N_B}. \tag{6·3}$$

When the value of q given in (2·1) is substituted in (6·1) and (6·2) Miller's formulae (29) and (30) of 1943 are obtained. In particular when $r=2$, $q=2(z-1)/z$, the formulae (821,10) and (821,11) of Fowler & Guggenheim (1939) are obtained. In making this substitution, it is convenient to note that from (2·1)

$$\frac{r-1}{r-q}=\tfrac{1}{2}z,\quad \frac{q-1}{r-q}=\tfrac{1}{2}z-1. \tag{6·4}$$

The simplest form of the vapour-pressure formulae is obtained from (6·1), (6·2) by using (6·4) where most convenient, but leaving q and r unchanged in other places. Thus

$$\frac{\lambda_A}{\lambda_{A,0}}=\frac{p_A}{p_{A,0}}=\frac{N_A}{(N_A+N_B/r)}\left(\frac{N_A+N_B/r}{N_A+N_B/q}\right)^{\frac{1}{2}zq} \tag{6·5}$$

$$\frac{\lambda_B}{\lambda_{B,0}}=\frac{p_B}{p_{B,0}}=\frac{N_B}{rN_A+N_B}\left(\frac{rN_A+N_B}{qN_A+N_B}\right)^{\frac{1}{2}z}. \tag{6·6}$$

In the special case of a linear array $z = 2$, $q = 1$ and formulae (6·5), (6·6) reduce to Raoult's law:

$$\frac{\lambda_A}{\lambda_{A,0}} = \frac{p_A}{p_{A,0}} = \frac{N_A}{N_A+N_B}, \tag{6·7}$$

$$\frac{\lambda_B}{\lambda_{B,0}} = \frac{p_B}{p_{B,0}} = \frac{N_B}{N_A+N_B}. \tag{6·8}$$

7. *The constant* ρ. It is noteworthy that it has not been necessary to evaluate ρ to determine how the partial vapour pressures depend on the composition of the mixture. It was sufficient to know that ρ is independent of N_A and N_B. It is nevertheless of some interest to determine ρ, since it is related to the number of possible ways of arranging N_A molecules A and N_B molecules B on $N_S = rN_A+N_B$ sites. This number, which is denoted by $g_{r,1}(N_A, N_B)$, is related to the thermodynamic functions by

$$-\frac{F}{kT} = -\frac{G}{kT} = \log g_{r,1}(N_A, N_B) + N_A \log f_A + N_B \log f_B. \tag{7·1}$$

Comparing (7·1) with (5·8), one finds,

$$\log g_{r,1}(N_A, N_B) = -\log N_A! - \log N_B!$$
$$+\frac{r-1}{r-q}\log(qN_A+N_B)! - \frac{q-1}{r-q}\log(rN_A+N_B)! + N_A\log\rho. \tag{7·2}$$

To determine the significance of ρ, consider the case $rN_A \ll N_B$ when, by neglecting $(r-q)N_A$ compared with $N_S = rN_A+N_B$, (7·2) becomes asymptotically equivalent to

$$\log g_{r,1}(N_A, N_B) = \log\frac{(rN_A+N_B)^{N_A}}{N_A!} + N_A\log\rho, \tag{7·3}$$

or

$$g_{r,1}(N_A, N_B) = \frac{(rN_A+N_B)^{N_A}}{N_A!}\rho^{N_A}. \tag{7·4}$$

Now provided $rN_A \ll N_B$, the factor multiplying ρ^{N_A} on the right of (7·4) is the number of ways of choosing N_A out of rN_A+N_B sites. It may therefore be regarded as the number of ways of choosing sites for one selected element in each of the N_A molecules. Consequently ρ must denote the number of distinct configurations of a molecule A when one of its elements is fixed on a given site.

Now that the physical significance of ρ is known, it is possible to specify its value in simple cases. Thus in the trivial case of molecules occupying a single site $r = 1$, $\rho = 1$. For molecules occupying two sites $\rho = z/\sigma$, where σ is the symmetry number of the molecule. Substituting z/σ for ρ in (7·2) one recovers formula (821,3) of Fowler & Guggenheim (1939), which in turn is equivalent to Chang's formula (33) of 1939.

For molecules occupying three sites, there are three distinct cases. If the molecules are linear and rigid $\rho = z/\sigma$. If the molecules are non-linear and rigid

$\rho = zz'/\sigma$, where z' is the number of alternative sites for the third element when the first two elements have been placed. If the molecules are flexible $\rho = z(z-1)/\sigma$.

For molecules occupying more than three sites there are numerous alternatives. The minimum value for ρ is z/σ for rigid straight molecules. The maximum value for ρ is approximately $z(z-1)^{r-2}/\sigma$ for entirely flexible molecules, but this is slightly inaccurate because it includes certain configurations in which the long molecule bends back on itself and two elements occupy the same site; such configurations should of course be excluded, but the error due to their inclusion is probably small. Some of these values of ρ are discussed by Miller (1942, p. 116), who also pointed out that a knowledge of ρ is not required for the determination of the dependence of the partial pressures on the composition of the mixture.

In the values assigned to ρ I have throughout included the factor $1/\sigma$ so as to conform with the usage adopted by Chang and Miller. It is, however, a matter of convention whether this factor $1/\sigma$ is included in f or in ρ. It must be inserted in one, but not in both.

8. *Generalization.* The method of approach described above is so simple that it can be applied without any difficulty to the most general system containing any number of different types of molecule each with its own geometric properties. Let the molecules of type i be N_i in number; let each such molecule occupy r_i sites; let the number of sites which are neighbours of a molecule of type i be $q_i z_i$; and let the number of alternative orientations of a molecule of type i be ρ_i when one of its elements has already been fixed. The q_i's and r_i's are related by

$$\tfrac{1}{2}z(r_i - q_i) = r_i - 1 \quad (\text{all } i). \tag{8·1}$$

Let the partition function of a molecule of type i, when attached to a prescribed set of sites be f_i. The procedure is so straightforward and so precisely analogous to that already used, that the steps can be left to the reader and only the results need be quoted. If $g(N_i)$ denotes the total number of possible arrangements of the molecules, then

$$\log g(N_i) = \tfrac{1}{2}z \log(\Sigma_i q_i N_i)! - \Sigma_i \log N_i! - (\tfrac{1}{2}z - 1)\log(\Sigma_i r_i N_i)! + \Sigma_i N_i \log \rho_i, \tag{8·2}$$

and

$$\frac{F}{kT} = \frac{G}{kT} = -\log g(N_i) - \Sigma_i N_i \log f_i. \tag{8·3}$$

From this by differentiation with respect to N_j one derives

$$\rho_j \lambda_j f_j = \frac{N_j(\Sigma_i r_i N_i)^{(\frac{1}{2}z-1)r_j}}{(\Sigma_i q_i N_i)^{\frac{1}{2}zq_j}} = \frac{N_j(\Sigma_i r_i N_i)^{\frac{1}{2}zq_j - 1}}{(\Sigma_i q_i N_i)^{\frac{1}{2}zq_j}}. \tag{8·4}$$

If in particular the molecules of type 1 occupy only one site, so that $r_1 = 1$, $q_1 = 1$, $\rho_1 = 1$, then

$$\lambda_1 f_1 = \frac{N_1(\Sigma_i r_i N_i)^{\frac{1}{2}z-1}}{(\Sigma_i q_i N_i)^{\frac{1}{2}z}}. \tag{8·5}$$

From (8·4) and (8·5) it follows that

$$\rho_j \frac{\lambda_j f_j}{(\lambda_1 f_1)^{r_j}} = \frac{N_j(\Sigma_i q_i N_i)^{\frac{1}{2}z(r_j - q_j)}}{N_1^{r_j}} = \frac{N_j(\Sigma_i q_i N_i)^{r_j - 1}}{N_1^{r_j}}, \tag{8·6}$$

using (8·1). The relation (8·6) is the extension of (4·6) and can be derived by the same elementary kinetic reasoning.

The general formula for the partial vapour pressures is

$$\frac{p_j}{p_{j,0}} = \frac{N_j(\Sigma_i r_i N_i/r_j)^{\frac{1}{2}zq_j-1}}{(\Sigma_i q_i N_i/q_j)^{\frac{1}{2}zq_j}}. \tag{8·7}$$

Formula (8·7) is the generalization of Raoult's law for molecules occupying various numbers of sites.

9. *Degree of accuracy.* As this paper is concerned only with systems with zero energies of mixing, there is no error due to the assumption of complete randomness of mixing. There is however a concealed error in my interpretation of complete randomness in the neighbourhood of a site occupied in a specified manner. Consider, for example, a site P neighbouring a site Q assumed to be occupied by an element of a molecule A. Then in the argument I have assumed that all conceivable manners of occupation of the site P, not prevented by the occupation of the site Q, are equally probable. Actually I have committed an error in not excluding those manners of occupation of the site P such that some element of the molecule occupying the site P is competing for some site R with some element of the molecule A assumed to be occupying the site Q. It is prohibitively difficult to assess the seriousness of this inaccuracy, but I believe it to be small.

It should be noted that the whole treatment applies strictly to a well-defined lattice, and it remains uncertain to what degree of accuracy the treatment may be applied to liquid mixtures. There seems, however, no reason to doubt that for liquid mixtures of molecules with the geometrical properties assumed formula (8·7) should be more accurate than Raoult's law.

I am indebted to Sir Ralph Fowler, F.R.S., for helpful and constructive criticism.

REFERENCES

Chang 1939*a* *Proc. Roy. Soc.* A, **169**, 512.
Chang 1939*b* *Proc. Camb. Phil. Soc.* **35**, 265.
Fowler & Guggenheim 1939 *Statistical Thermodynamics.* Cambridge University Press.
Fowler & Rushbrooke 1937 *Trans. Faraday Soc.* **33**, 1272.
Guggenheim 1932 *Proc. Roy. Soc.* A, **135**, 181.
Guggenheim 1933 *Modern Thermodynamics*, p. 103. Methuen.
Guggenheim 1937 *Trans. Faraday Soc.* **33**, 151.
Hildebrand 1924 *Solubility*, p. 26. Chemical Catalog Co.
Lewis 1908 *J. Amer. Chem. Soc.* **30**, 668.
Miller 1942 *Proc. Camb. Phil. Soc.* **38**, 109.
Miller 1943 *Proc. Camb. Phil. Soc.* **39**, 54.
Van't Hoff 1903 *Vorlesungen über Theoretische und Physikalische Chemie*, Part 2.
Washburn 1910 *Z. phys. Chem.* **74**, 537.

Statistical thermodynamics of mixtures with non-zero energies of mixing

By E. A. Guggenheim, M.A., Sc.D., *Imperial College of Science.*

(*Communicated by Sir Ralph Fowler, F.R.S.—Received* 5 *April* 1944)

A combinatory formula is obtained for $g(N_i, X_{ij})$, the number of ways of arranging a mixture of any number of kinds of molecules on a lattice, the values of N_i and X_{ij} being specified, where N_i denotes the number of molecules of type i, z denotes the number of sites which are neighbours of one site, and zX_{ij} denotes the number of pairs of neighbouring sites occupied one by a molecule of type i, the other by a molecule of type j. Each molecule of type i is assumed to occupy r_i sites, where r_i is any integer with different values for different types of molecules.

This formula is used to derive the thermodynamic properties of mixtures of molecules occupying various numbers of sites, assuming that the intermolecular energy can be regarded as a sum of terms, each pair of neighbours contributing one term.

For binary mixtures the formulae obtained are very similar to those previously obtained for 'regular' solutions where each molecule occupies one site. A rather simple formula is obtained for the critical temperature and the composition of the critical mixture.

The degree of accuracy of the treatment is the same as Chang's use of Bethe's first approximation and as the 'quasi-chemical' method of approach. A brief investigation of a higher approximation for a binary regular mixture on a close-packed lattice indicates that the errors due to the approximation used are unlikely ever to be serious.

1. *Introduction.* In a former paper (1944) I derived a general formula for $g(N_i)$, the total number of distinguishable ways of arranging on a lattice a mixture of any number of kinds of molecules differing amongst themselves in the number of sites occupied by one molecule, the number of molecules of type i being N_i. From this I deduced the thermodynamic properties of such a mixture subject to the restriction that all the configurations had the same energy, or in other words that the energies of mixing were all zero.

The thermodynamic properties of such a mixture having non-zero energies of mixing are not determinable from $g(N_i)$, but from the more complicated quantity $g(N_i, W)$ denoting the number of distinguishable configurations of intermolecular energy W. The configurational part of the free energy is equal to

$$-kT\log\Sigma\{g(N_i, W)\exp(-W/kT)\},$$

where the summation extends over all values of W. As usual, for macroscopic systems there is no sensible loss of accuracy if the sum $\Sigma\{\ \}$ is replaced by its maximum term, say $g(N_i, \overline{W})\exp(-\overline{W}/kT)$, where $\overline{W}$ is the value of W which maximizes $g(N_i, W)\exp(-W/kT)$ and is the configurational thermodynamic energy of the system in equilibrium.

The evaluation of $g(N_i, W)$ by direct methods is prohibitively difficult, but the following considerations are sufficient to determine unambiguously the form of $g(N_i, W)$.

In the first place $\{\log g(N_i, W)\}/\Sigma_i N_i$ must be homogeneous of degree zero in the N_i's, so as to ensure that the free energy of a system of given temperature and composition shall be proportional to the size of the system.

In the second place when W is zero $\log g(N_i, \overline{W})$ must, apart from terms negligible compared with N_i, reduce to $\log g(N_i)$ so as to ensure that the free energy is a continuous function of W in the neighbourhood of $W = 0$.

In the third place I shall assume that the conditions for maximizing

$$g(N_i, W)\exp(-W/kT)$$

with respect to W are equivalent to conditions of a 'quasi-chemical equilibrium' (Guggenheim 1938) between the alternative methods of occupation of a pair of neighbouring sites. This is an assumption which has proved fruitful in the simple case where each molecule of whatever type occupies a single site (Guggenheim 1935; Rushbrooke 1938). In order not to interrupt the argument this assumption will be used at an early stage, discussion of its significance being postponed to §§ 12–14.

Having obtained a formula for $g(N_i, W)$ and having maximized

$$g(N_i, W)\exp(-W/kT)$$

with respect to W, the thermodynamic properties of the system are formally determined. The equations of quasi-chemical equilibrium for maximizing with respect to $W \neq 0$ can, however, be solved explicitly only for binary mixtures. For these the thermodynamic functions can be expressed in a form having a close resemblance to those previously obtained for 'regular' binary mixtures of molecules each occupying a single site. The formulae for binary mixtures include as a special case those given recently by Orr (1944) for a mixture of two kinds of molecule, of which those of one kind occupy only one site each. The present work was in fact stimulated by Orr's paper, which I saw before its publication.

2. *Notation and counting operations.* Let N_i denote the number of molecules of type i. Let each molecule of type i occupy r_i sites. Let the number of sites which are neighbours of any one site be z. It will be assumed that the structures of all the molecules are such that the elements which occupy a single site are interconnected in the molecule as a simple chain or a branched chain, but that there are no closed rings in the molecule. It is convenient to define parameters q_i, one for each molecular type, by the relation

$$z(r_i - q_i) = 2(r_i - 1) \quad (\text{all } i). \tag{2·1}$$

The physical measuring of q_i is that $q_i z$ is the number of sites neighbours of the r_i sites occupied by a molecule of type i, excluding those neighbours of each site occupied by the next element of the same molecule. Apart from configurations of a long molecule bent back on itself so that 'distant' elements of the same molecule occupy neighbouring sites, the definition of q_i could be simplified as follows: $q_i z$ is the number of sites which are neighbours of a molecule of type i.

Let f_i denote the partition function, omitting symmetry numbers, of a molecule of type i attached to a given set of sites, f_i being assumed independent of this set of sites. This assumption is accurate provided each possible internal rotation in a molecule is either unrestricted or completely restricted.

Let σ_i denote the symmetry number of a molecule of type i. Let ρ_i denote the number of alternative configurations of a molecule of type i when one of its elements is fixed on a given site. When $r_i = 1$ the value of ρ_i is 1. When $r_i \geqslant 2$ the minimum value of ρ_i is z/σ_i for rigid straight molecules. The maximum value of ρ_i is nearly $z(z-1)^{r-2}/\sigma_i$ for completely flexible molecules. The last mentioned value of ρ_i is actually somewhat greater than the maximum possible because it includes certain configurations in which a molecule is bent back on itself so that two 'distant' elements occupy the same site; such configurations should of course be excluded.

It is assumed that in any given configuration of the system the intermolecular potential energy W consists of a sum of terms, each pair of neighbouring sites occupied by elements of different molecules contributing one term. Let w_{ij} denote the contribution to the potential energy by a pair of sites of which one is occupied by an element of a molecule of type i, the other by an element of a different molecule of type j. In order to achieve brevity in the formulae the energy zeros are so chosen that $w_{ii} = 0$ for all i. If it is wished to remove this restriction, it is merely necessary to replace w_{ij} in all formulae by $w_{ij} - \frac{1}{2}w_{ii} - \frac{1}{2}w_{jj}$.

Let the number of pairs of neighbouring sites occupied by different molecules of which one is of type i, the other of type j be denoted by zX_{ij}. Then according to these definitions

$$2zX_{ii} + \Sigma'_k zX_{ik} = q_i zN_i, \tag{2·2}$$

where Σ'_k denotes summation over all $k \neq i$. A more convenient form of (2·2) is

$$2X_{ii} = q_i N_i - \Sigma'_k X_{ik}. \tag{2·3}$$

The total intermolecular potential energy W of the system is then given by

$$W = z\Sigma'_{ij} X_{ij} w_{ij}, \tag{2·4}$$

where Σ'_{ij} denotes summation over all distinct pairs of types of molecules.

3. *Quasi-chemical equilibrium.* If one uses the symbol (ij) to denote a pair of sites occupied by elements of two different molecules, one of type i the other of type j, and considers a shuffling of the molecules such that two (ij)'s are created and an (ii) and a (jj) are destroyed, one can write this symbolically

$$(ii) + (jj) \rightarrow 2(ij). \tag{3·1}$$

Formula (3·1) recalls the formula for a chemical reaction. It is not unreasonable to expect that the equilibrium condition between (ii), (jj) and (ij) should take a form similar to that corresponding to a chemical reaction. If (3·1) represented a chemical reaction between molecules, the equilibrium condition would be

$$\frac{X_{ij}^2}{4X_{ii}X_{jj}} = e^{-2w_{ij}/kT} \quad (i \neq j), \tag{3·2}$$

where the molecular partition functions of ii, jj and ij are assumed to be identical except for symmetry factors of $\frac{1}{2}$ in the first two. If a formula such as (3·2) is applied to the number of pairs of neighbouring sites occupied in several ways it is conveniently referred to as a condition of 'quasi-chemical' equilibrium.

An equation of quasi-chemical equilibrium similar to (3·2) was first proposed for 'regular' mixtures of molecules each occupying a single site (Guggenheim 1935). An error in this first use of the equation was pointed out by Rushbrooke (1938), who at the same time showed that the equation could be derived from the approximate formulation of grand partition functions usually referred to as Bethe's method It was later shown that the equivalence between the equations of quasi-chemical equilibrium and those of Bethe's method were quite general for molecules or atoms each occupying a single site, not only for all 'regular' systems (Guggenheim 1938), but also for systems with long-range order (Fowler & Guggenheim 1940). It is therefore not surprising to find that the two methods are also equivalent for systems of molecules each occupying several sites. To avoid a long digression in the reasoning, discussion of this equivalence is postponed to a later stage in § 12.

It is convenient to substitute from (2·3) and a similar formula for X_{jj} into (3·2), thus obtaining for the conditions of quasi-chemical equilibrium

$$\frac{X}{(q_iN_i-\Sigma'_kX_{ik})\,(q_jN_j-\Sigma'_kX_{jk})}=e^{-2w_{ij}/kT}\quad(i\neq j),. \tag{3·3}$$

If one denotes by X^*_{ij} the equilibrium values of X_{ij} in the particular case that all energies of mixing are zero, that is to say all $w_{ij}=0$, then

$$X^{*2}_{ij}=(q_iN_i-\Sigma'_kX^*_{ik})\,(q_jN_j-\Sigma'_kX^*_{jk})\quad(i\neq j). \tag{3·4}$$

The relevant solution of (3·4) is

$$X^*_{ij}=\frac{q_iN_iq_jN_j}{\Sigma_kq_kN_k}\quad(i\neq j), \tag{3·5}$$

$$2X^*_{ii}=\frac{(q_iN_i)^2}{\Sigma_kq_kN_k}. \tag{3·6}$$

4. *Combinatory factor* $g(N_i)$. The total number $g(N_i)$ of distinguishable ways of arranging all the molecules, N_i of each species i, on $\Sigma_ir_iN_i$ sites has been shown in the previous paper (1944) to be given by

$$\log g(N_i)=\tfrac{1}{2}z\log(\Sigma_iq_iN_i)!-(\tfrac{1}{2}z-1)\log(\Sigma_ir_iN_i)!-\Sigma_i\log N_i!+\Sigma_iN_i\log\rho_i, \tag{4·1}$$

where each q_i is related to the corresponding r_i by (2·1).

5. *Combinatory factor* $g(N_i, X_{ij})$. It is now possible to construct an expression for the number of distinguishable configurations of all the molecules having a total intermolecular energy W. In the introduction I called this number $g(N_i, W)$, but W is related to the X_{ij}'s according to (2·4), and it will be more convenient to consider this number as a function of the N_i's and the X_{ij}'s. I shall accordingly henceforth call it $g(N_i, X_{ij})$.

The formula for $g(N_i, X_{ij})$ determined by the considerations enumerated in § 1 is

$$\begin{aligned}\log g(N_i, X_{ij}) &= \log g(N_i) + \Sigma_i \log \frac{\{zX^*_{ii}\}!}{(zX_{ii})!} + \Sigma'_{ij} \log \frac{\{zX^*_{ij}\}!\, 2^{zX_{ij}}}{\{zX_{ij}\}!\, 2^{zX^*_{ij}}} \\ &= \log g(N_i) + \Sigma_i \log \frac{\{\tfrac{1}{2}z(q_iN_i - \Sigma'_k X^*_{ik})\}!}{\{\tfrac{1}{2}z(q_iN_i - \Sigma'_k X_{ik})\}!} + \Sigma'_{ij} \log \frac{\{zX^*_{ij}\}!\, 2^{zX_{ij}}}{\{zX_{ij}\}!\, 2^{zX^*_{ij}}}, \qquad (5\cdot 1)\end{aligned}$$

where $g(N_i)$ is given by (4·1). It can now be verified that $g(N_i, X_{ij})$ as defined by (5·1) does in fact have the properties enumerated in § 1.

In the first place if one maximizes

$$\log\{g(N_i, X_{ij})\, e^{-W/kT}\} = \log\{g(N_i, X_{ij})\, e^{-z\Sigma'_{ik}X_{ik}w_{ik}/kT}\} \qquad (5\cdot 2)$$

with respect to X_{ij}, the condition obtained is

$$\tfrac{1}{2}z \log(q_iN_i - \Sigma'_k X_{ik}) + \tfrac{1}{2}z \log(q_jN_j - \Sigma'_k X_{jk}) - z\log X_{ij} - zw_{ij}/kT = 0 \quad (i \neq j), \qquad (5\cdot 3)$$

or

$$\frac{X^2_{ij}}{(q_iN_i - \Sigma'_k X_{ik})(q_jN_j - \Sigma'_k X_{jk})} = e^{-2w_{ij}/kT} \quad (i \neq j), \qquad (5\cdot 4)$$

which is identical with (3·3). This confirms that the maximizing conditions when $g(N_i, X_{ij})$ is given by (5·1) are of the required quasi-chemical form.

In the second place it is seen from (5·4) that the X_{ij}'s are homogeneous of the first degree in the N_i's, from which it follows that $\{\log g(N_i, X_{ij})\}/\Sigma_i N_i$ as given by (5·1) is homogeneous of zero degree in the N_i's as required.

Finally, from the definition of the X^*_{ij}'s it is evident that in the special case that all $w_{ij} = 0$, that is to say all energies of mixing zero, $g(N_i, \bar{X}_{ij})$ reduces to $g(N_i)$. This completes the justification of formula (5·1) for $g(N_i, X_{ij})$.

When the values of X^*_{ij} given by (3·5) and the value of $g(N_i)$ given by (4·1) are substituted into (5·1), this can be rewritten as

$$\begin{aligned}\log g(N_i, X_{ij}) &= \Sigma_i N_i \log \rho_i - \Sigma_i \log N_i! + z\Sigma_i \log(q_iN_i)! \\ &\quad - (\tfrac{1}{2}z - 1)\log(\Sigma_i r_i N_i)! - \tfrac{1}{2}z\Sigma_i \log(q_iN_i - \Sigma'_k X_{ik})! - z\Sigma'_{ij}\log X_{ij}!. \qquad (5\cdot 5)\end{aligned}$$

It is worthy of notice that the term $\tfrac{1}{2}z\log(\Sigma_i q_i N_i)!$ in $\log g(N_i)$ is cancelled by an equal but opposite term coming from the denominators of (3·5) and (3·6). Consequently $g(N_i, X_{ij})$ contains no term in $\log(\Sigma_i q_i N_i)!$. Formula (5·5) is the fundamental one in the present treatment.

6. *Thermodynamic functions.* The free energy of the system is, at negligible pressures, related to $g(N_i, X_{ij})$ by

$$\begin{aligned}F &= -kT\Sigma_i N_i \log f_i - kT\log g(N_i, \bar{X}_{ij}) + \bar{W} \\ &= -kT\Sigma_i N_i \log f_i - kT\log g(N_i, \bar{X}_{ij}) + z\Sigma'_{ij}\bar{X}_{ij}w_{ij}, \qquad (6\cdot 1)\end{aligned}$$

where $\bar{X}_{ij}$ denotes the equilibrium value of X_{ij} determined by equations (3·3). Substituting from (5·5) into (6·1), one obtains

$$\begin{aligned}F/kT &= -\Sigma_i N_i \log f_i\rho_i + \Sigma_i \log N_i! - z\Sigma_i \log(q_iN_i)! + (\tfrac{1}{2}z - 1)\log(\Sigma_i r_i N_i)! \\ &\quad + \tfrac{1}{2}z\Sigma_i \log(q_iN_i - \Sigma'_k \bar{X}_{ik})! + z\Sigma'_{ij}\log \bar{X}_{ij}! + z\Sigma'_{ij}\bar{X}_{ij}w_{ij}. \qquad (6\cdot 2)\end{aligned}$$

The partial potentials μ_i and the absolute activities λ_i (Fowler & Guggenheim 1939, § 224) can be obtained by differentiating (6·2) with respect to N_i, and in so doing variations in $\bar{X}_{ij}$ may be ignored since it is known that $\partial F/\partial X_{ij} = 0$ for all i, j. It is thus found that

$$\begin{aligned}\mu_i/kT = \log \lambda_i = &-\log f_i \rho_i + \log N_i - zq_i \log q_i N_i \\ &+ (\tfrac{1}{2}z - 1)\, r_i \log \Sigma_j r_j N_j + \tfrac{1}{2}zq_i \log (q_i N_i - \Sigma'_k \bar{X}_{ik}),\end{aligned} \tag{6·3}$$

or

$$\lambda_i = \frac{N_i}{f_i \rho_i} \frac{(\Sigma_j r_j N_j)^{(\frac{1}{2}z-1)r_i} (q_i N_i - \Sigma'_k \bar{X}_{ik})^{\frac{1}{2}zq_i}}{(q_i N_i)^{zq_i}}. \tag{6·4}$$

This can be rewritten as

$$\lambda_i = \frac{N_i}{f_i \rho_i} \frac{(\Sigma_j r_j N_j)^{(\frac{1}{2}z-1)r_i} (2\bar{X}_{ii})^{\frac{1}{2}zq_i}}{(q_i N_i)^{zq_i}}, \tag{6·5}$$

by using (2·3). Finally, by using (2·1) the alternative formula

$$\lambda_i = \frac{N_i}{f_i \rho_i \Sigma_j r_j N_j} \left\{ \frac{(\Sigma_j r_j N_j)\, 2\bar{X}_{ii}}{(q_i N_i)^2} \right\}^{\frac{1}{2}zq_i} \tag{6·6}$$

is obtained.

In the particular case of all energies of mixing zero, $\bar{X}_{ii}$ becomes X^*_{ii}, and using (3·6) in (6·6) one obtains

$$\lambda_i = \frac{N_i}{f_i \rho_i \Sigma_j r_j N_j} \left\{ \frac{\Sigma_j r_j N_j}{\Sigma_j q_j N_j} \right\}^{\frac{1}{2}zq_i}. \tag{6·7}$$

which is formula (8·4) of my previous paper (1944*b*).

7. *Partial vapour pressures.* From (6·6) one deduces the relation between the partial vapour pressure p_i of the component i in the mixture and the vapour pressure $p_{i,0}$ of the pure substance

$$\frac{p_i}{p_{i,0}} = \frac{r_i N_i}{\Sigma_j r_j N_j} \left\{ \frac{q_i(\Sigma_j r_j N_j)\, 2\bar{X}_{ii}}{r_i (q_i N_i)^2} \right\}^{\frac{1}{2}zq_i} = \frac{r_i N_i}{\Sigma_j r_j N_j} \left\{ \frac{q_i(\Sigma_j r_j N_j)\,(q_i N_i - \Sigma'_k \bar{X}_{ik})}{r_i (q_i N_i)^2} \right\}^{\frac{1}{2}zq_i}. \tag{7·1}$$

In this formula and in the formulae of the preceding section the $\bar{X}_{ik}$'s have the values determined by the quasi-chemical equations (3·3). Since, however, these equations can generally not be solved explicitly, no further simplification is possible except in the case of a binary mixture, which will be considered in greater detail.

8. *Combinatory formula for binary mixtures.* When there are only two types of molecules, say A and B, in the mixture, formula (5·5) becomes

$$g(N_A, N_B, X) = \frac{\rho_A^{N_A} \rho_B^{N_B} \{(q_A N_A)!\}^z \{(q_B N_B)!\}^z}{N_A!\, N_B!\, \{(r_A N_A + r_B N_B)!\}^{\frac{1}{2}z-1} \{(q_A N_A - X)!\}^{\frac{1}{2}z} \{X!\}^z \{(q_B N_B - X)!\}^{\frac{1}{2}z}}, \tag{8·1}$$

where X has been written for X_{AB}.

Formulae equivalent to two special cases of (8·1) have previously been obtained by Chang (1939). The first case was the simplest possible one,

$$r_A = r_B = q_A = q_B = \rho_A = \rho_B = 1.$$

The second case treated by Chang was

$$r_A = 2, \quad q_A = 2(z-1)/z, \quad \rho_A = z/2, \quad r_B = q_B = \rho_B = 1.$$

In the case of a binary mixture there is only one equation of quasi-chemical equilibrium, namely,

$$X^2 = (q_A N_A - X)(q_B N_B - X)\, e^{-2w/kT}, \tag{8·2}$$

where w has been written for w_{AB}. It is convenient to define a fraction x by

$$x = \frac{q_A N_A}{q_A N_A + q_B N_B}, \quad 1 - x = \frac{q_B N_B}{q_A N_A + q_B N_B}, \tag{8·3}$$

and a quantity β by

$$\beta = \{(1-2x)^2 + 4x(1-x)\, e^{2w/kT}\}^{\frac{1}{2}}. \tag{8·4}$$

It is to be noted in the simplest case of 'regular' solutions $q_A = q_B = 1$ and x is the mole-fraction of A.

The solution $\overline{X}$ of the quadratic equation (8·2) is conveniently written in the form

$$\frac{\overline{X}}{q_A N_A + q_B N_B} = \frac{2x(1-x)}{\beta + 1}, \tag{8·5}$$

from which follow

$$\frac{q_A N_A - \overline{X}}{q_A N_A + q_B N_B} = \frac{x(\beta - 1 + 2x)}{\beta + 1}, \tag{8·6}$$

$$\frac{q_B N_B - \overline{X}}{q_A N_A + q_B N_B} = \frac{(1-x)(\beta + 1 - 2x)}{\beta + 1}. \tag{8·7}$$

9. *Thermodynamic functions for binary mixtures.* The neatest, if not the most direct, way to obtain a compact formula for the free energy is via the absolute activities λ_A and λ_B. For a binary mixture formula (6·6) becomes

$$\lambda_A = \frac{N_A}{f_A \rho_A (r_A N_A + r_B N_B)} \left\{\frac{(r_A N_A + r_B N_B)(q_A N_A - \overline{X})}{(q_A N_A)^2}\right\}^{\frac{1}{2} z q_A}, \tag{9·1}$$

$$\lambda_B = \frac{N_B}{f_B \rho_B (r_A N_A + r_B N_B)} \left\{\frac{(r_A N_A + r_B N_B)(q_B N_B - \overline{X})}{(q_B N_B)^2}\right\}^{\frac{1}{2} z q_B}. \tag{9·2}$$

Substituting from (8·6) and (8·7) into (9·1) and (9·2) respectively, one obtains

$$\lambda_A = \frac{N_A}{f_A \rho_A (r_A N_A + r_B N_B)} \left\{\frac{(r_A N_A + r_B N_B)(\beta - 1 + 2x)}{(q_A N_A + q_B N_B)(\beta + 1)\, x}\right\}^{\frac{1}{2} z q_A}, \tag{9·3}$$

$$\lambda_B = \frac{N_B}{f_B \rho_B (r_A N_A + r_B N_B)} \left\{\frac{(r_A N_A + r_B N_B)(\beta + 1 - 2x)}{(q_A N_A + q_B N_B)(\beta + 1)(1 - x)}\right\}^{\frac{1}{2} z q_B}. \tag{9·4}$$

If the subscript 0 is used to denote the value of a quantity for the substance in the pure state, (9·3) and (9·4) can be rewritten as

$$\frac{\lambda_A}{\lambda_{A,0}} = \frac{p_A}{p_{A,0}} = \frac{N_A}{(N_A + r_B N_B / r_A)} \left\{\frac{(N_A + r_B N_B / r_A)(\beta - 1 + 2x)}{(N_A + q_B N_B / q_B)(\beta + 1)\, x}\right\}^{\frac{1}{2} z q_A}, \tag{9·5}$$

$$\frac{\lambda_B}{\lambda_{B,0}} = \frac{p_B}{p_{B,0}} = \frac{N_B}{(r_A N_A / r_B + N_B)} \left\{\frac{(r_A N_A / r_B + N_B)(\beta + 1 - 2x)}{(q_A N_A / q_B + N_B)(\beta + 1)(1 - x)}\right\}^{\frac{1}{2} z q_B}. \tag{9·6}$$

In the simplest case $r_A = r_B = q_A = q_B = 1$, these formulae reduce to those given by Fowler & Guggenheim (1939, § 819). The formulae for the case $r_B = q_B = 1$, $r_A > 1$ have been derived in a recent paper by Orr (1944).

The free energy at negligible pressures is obtained conveniently from the thermodynamic relation

$$F/kT = N_A \log \lambda_A + N_B \log \lambda_B. \tag{9·7}$$

The result can be written

$$F = F_1 + F_2, \tag{9·8}$$

where

$$\begin{aligned} F_1/kT &= -\log g(N_A, N_B) - N_A \log f_A - N_B \log f_B \\ &= -N_A \log f_A \rho_A - N_B \log f_B \rho_B + \log N_A! + \log N_B! \\ &\quad - \tfrac{1}{2}z \log (q_A N_A + q_B N_B)! + (\tfrac{1}{2}z - 1) \log (r_A N_A + r_B N_B)! \end{aligned} \tag{9·9}$$

and

$$F_2/kT = \tfrac{1}{2}z(q_A N_A + q_B N_B) \left\{ x \log \frac{\beta - 1 + 2x}{x(\beta + 1)} + (1 - x) \log \frac{\beta + 1 - 2x}{(1 - x)(\beta + 1)} \right\}. \tag{9·10}$$

In the case of zero energy of mixing $\beta = 1$ and F_2 vanishes, whereas F_1 is independent of the energy of mixing. Thus F_2 may be regarded as the part of the free energy due to the energy of mixing. The same formula for the free energy can be obtained, though more laboriously by using (8·1) and (8·5) in

$$F/kT = -N_A \log f_A - N_B \log f_B - \log g(N_A, N_B, \overline{X}) + z\overline{X}w/kT. \tag{9·11}$$

10. *Energy of mixing*. Having obtained formulae for the free energy, one can derive the energy according to the usual thermodynamic formulae by differentiation of F/T with respect to $1/T$. The only terms in F_1/T which depend on the temperature are $-N_A \log f_A - N_B \log f_B$, and the contribution from these terms to the energy remains unchanged if the mixture is separated into its components. Hence the energy of mixing E is due entirely to F_2/T and is given by

$$E = \frac{1}{T} \frac{dF_2}{d\beta} \frac{d\beta}{d(1/T)}. \tag{10·1}$$

By straightforward, if lengthy, algebra it is found that

$$\frac{dF_2}{d\beta} = \tfrac{1}{2}zkT(q_A N_A + q_B N_B) \frac{4x(1 - x)\beta}{\{\beta^2 - (1 - 2x)^2\}\{\beta + 1\}}, \tag{10·2}$$

$$\frac{1}{T} \frac{d\beta}{d(1/T)} = \frac{w}{kT} \frac{\beta^2 - (1 - 2x)^2}{\beta}, \tag{10·3}$$

whence

$$E = zw(q_A N_A + q_B N_B) \frac{2x(1 - x)}{\beta + 1}. \tag{10·4}$$

Comparing (10·4) with (8·5), one finds

$$E = z\overline{X}w = \overline{W}, \tag{10·5}$$

as might be expected. This circuitous derivation of the physically obvious (10·5) is a useful check on the self-consistency of the formulae.

11. *Critical mixing.* A necessary condition for the internal stability of a mixture is that

$$\frac{1}{kT}\frac{\partial \mu_A}{\partial x} = \frac{\partial \log \lambda_A}{\partial x} > 0. \tag{11·1}$$

If at a given temperature this inequality is not satisfied by all mixtures, then some mixtures will be unstable and will split into two phases of different composition.

Differentiating (9·3) with respect to x, using (8·4), one obtains an expression which eventually reduces to

$$\frac{\partial \log \lambda_A}{\partial x} = \frac{1}{2x}\left\{\frac{zq_A}{\beta} - \frac{(z-2)\,r_A r_B/q_B}{r_A x/q_A + r_B(1-x)/q_B}\right\}. \tag{11·2}$$

For the sake of brevity two new constants a and b are now defined by

$$a = \frac{zq_A}{(z-2)\,r_A} = \frac{zq_A}{zq_A-2}, \tag{11·3}$$

$$b = \frac{zq_B}{(z-2)\,r_B} = \frac{zq_B}{zq_B-2}. \tag{11·4}$$

From (11·2) it follows that the equation

$$\partial \log \lambda_A/\partial x = 0 \tag{11·5}$$

reduces to

$$\beta = bx + a(1-x). \tag{11·6}$$

Combining (11·6) with (8·4), it is further found that (11·5) is equivalent to

$$(b^2-1)\,x^2 - 2(2e^{2w/kT} - 1 - ab)\,x(1-x) + (a^2-1)\,(1-x)^2 = 0. \tag{11·7}$$

At any given temperature some mixtures will be internally unstable if (11·7), regarded as a quadratic in x or in $x/(1-x)$, has real roots. The condition of critical mixing obtains when (11·7) has two equal roots. Hence using the subscript c to denote critical values, it follows that

$$e^{2w/kT_c} = \tfrac{1}{2}\{1 + ab + \sqrt{[(a^2-1)\,(b^2-1)]}\}, \tag{11·8}$$

$$\frac{x_c}{1-x_c} = \sqrt{\frac{a^2-1}{b^2-1}}. \tag{11·9}$$

By using the definitions (8·3) of x, (11·3) of a and (11·4) of b, (11·9) can be rewritten as

$$\left(\frac{N_A}{N_B}\right)_c = \frac{r_B q_B}{r_A q_A}\sqrt{\frac{q_A z-1}{q_B z-1}}. \tag{11·10}$$

The simplest example of these formulae corresponding to $r_A = r_B = q_A = q_B = 1$, $a = b = z/(z-2)$ was obtained by Rushbrooke (1938). The next simplest example corresponding to $r_A = 2$, $q_A = 2(z-1)/z$, $r_B = q_B = 1$, $a = (z-1)/(z-2)$, $b = z/(z-2)$ was obtained by Chang (1939).

12. *Quasi-chemical equilibrium and Bethe's method.* The treatment generally known as Bethe's method consists in the construction of an approximate grand

partition function for one (or sometimes two) central sites and their neighbours. The equivalence between Bethe's method and the quasi-chemical equations has been proved for molecules each occupying a single site (Rushbrooke 1938; Guggenheim 1938; Fowler & Guggenheim 1940). I shall now show that for the most general kind of mixture of molecules each occupying any number of sites Bethe's method always leads to the quasi-chemical conditions of equilibrium. More precisely I shall show that these quasi-chemical conditions are implicitly assumed in Bethe's method. I shall give a detailed proof only for occupation of sites by either A or B molecules. The proof for the general case is precisely similar, but requires a more elaborate notation.

In Bethe's method an approximate grand partition function is constructed for one (or two) central sites and their neighbours. There is one term in this grand partition function for each distinct manner of occupation of all the sites. Let (A, A, S) denote the term in the grand partition function corresponding to a selected central site being occupied by an element of an A molecule, a selected nearest neighbour being occupied by an element of a *different* A molecule, and all the remaining sites central and neighbours being occupied in some specified manner denoted symbolically by S. Let (B, B, S) be similarly defined. Let (A, B, S) denote the term corresponding to the selected central site being occupied by an element of an A molecule, the selected neighbour site by an element of a B molecule, all the remaining sites, central and neighbours being occupied in the manner S. Let (B, A, S) be defined similarly and correspond to the converse manner of occupation of the selected central and neighbour site.

One of the essential approximations of Bethe's method is the assumption

$$\left.\begin{aligned}(A,\ A,\ S)/(A,\ B,\ S)\ \text{independent of}\ S,\\ (B,\ A,\ S)/(B,\ B,\ S)\ \text{independent of}\ S.\end{aligned}\right\} \tag{12·1}$$

According to this assumption the two Bethe parameters ϵ and η can be defined by

$$(A, A, S)/(A, B, S) = \epsilon/\eta, \tag{12·2}$$

$$(B, A, S)/(B, B, S) = \epsilon\eta. \tag{12·3}$$

It is usual to obtain further information concerning the relation between ϵ and η by using the fact that the central site and its neighbour are physically equivalent. The following alternative procedure is simpler and at least as profitable.

Sum (12·2) and (12·3) over all S obtaining

$$(A, A)/(A, B) = \epsilon/\eta, \tag{12·4}$$

$$(B, A)/(B, B) = \epsilon\eta, \tag{12·5}$$

where (A, A) denotes the sum of all terms in the grand partition function corresponding to the selected central site being occupied by an element of an A molecule and the selected neighbour site by an element of a *different* A molecule, the (A, B),

(B, A) and (B, B) are similarly defined. Now divide (12·5) by (12·4), thus eliminating ϵ and obtaining

$$\frac{(A,B)(B,A)}{(A,A)(B,B)} = \eta^2. \tag{12·6}$$

Since the central site and its neighbour are physically equivalent, it is obvious that

$$(A,B) = (B,A), \tag{12·7}$$

and consequently (12·6) is equivalent to

$$\frac{\{(A,B)+(B,A)\}^2}{4(A,A)(B,B)} = \eta^2. \tag{12·8}$$

Now each term in the grand partition function is proportional to the frequency of occurrence of the manner of occupation to which that term refers. Moreover, there is no special significance attaching to the particular pair of sites selected as central site and its neighbour. The numbers of pairs of sites occupied both by elements of different A molecules, both by elements of different B molecules and one by an element of an A molecule, the other by an element of a B molecule are therefore proportional respectively to (AA), (BB) and $\{(A,B)+(B,A)\}$. Hence (12·8) is equivalent to the condition (3·2) of quasi-chemical equilibrium, provided

$$\eta = e^{-w_{AB}/kT}, \tag{12·9}$$

which is precisely the value *assumed* for η in Bethe's method.

To sum up, the two essential approximate assumptions of Bethe's method are (12·1) and (12·9). The two assumptions together lead immediately to the quasi-chemical conditions of equilibrium.

13. *Nature of approximation.* Neither the quasi-chemical method nor Bethe's method is completely accurate for the following reason. In both methods it is assumed that having specified the manner of occupation of one site, say a, then the probabilities of occupation of one of its neighbours, say b, in alternate manners are independent of those for another neighbour site say b'. Actually the manner of occupation of b affects the probabilities of its neighbours, say $c, c', \ldots$, and these affect their neighbour and so on. But b is related to b' not only through a, but also through closed rings $bcd\ldots b$. Hence the probabilities of occupation of b and b' are not truly independent.

The error due to this inexact assumption will be the more serious the shorter the rings connecting b to b' without passing through a. The error may be expected to be worst in a close-packed structure, in which b and b' can be neighbours of each other. An investigation of this worst case in the following section indicates that the inaccuracy is never serious.

14. *Higher approximation for close-packed lattice.* An estimate of the degree of inaccuracy of our formulae in the case of a close-packed lattice can be obtained by applying equations of the quasi-chemical type to sets of three sites forming triangles

and comparing the result with that already obtained by its application to pairs of sites. I shall consider only the simplest case of a binary mixture of molecules A and B each occupying a single site.

As previously, let z denote the number of neighbours of a single site. Let z' denote the number of triangles of three touching sites with any given pair of sites in common. Then the total number of triangles of three touching sites is $(N_A+N_B)zz'/6$. The numbers of triangles occupied in alternative manners now require for specification two independent quantities instead of X in the case of pairs of sites. Two such quantities Y and Y' can be defined such that the number of triangles occupied in each manner are as given in the following table:

AAA	$(N_A-2Y-Y')zz'/6$
AAB	$3Yzz'/6$
ABB	$3Y'zz'/6$
BBB	$(N_B-Y-2Y')zz'/6.$

Furthermore, Y and Y' are related to X, defined in § 2, by

$$X = Y+Y'. \tag{14·1}$$

The equations of quasi-chemical equilibrium for the triangles are

$$Y'(N_A-2Y-Y') = Y^2e^{2w/kT}, \tag{14·2}$$

$$Y(N_B-Y-2Y') = Y'^2e^{2w/kT}. \tag{14·3}$$

These equations are generally intractable, but they simplify considerably for a mixture of equal numbers of A and B molecules. Then the solution obtained is

$$\overline{Y} = \overline{Y}' = N_A/(3+e^{2w/kT}). \tag{14·4}$$

Denoting, as previously, by an asterisk the value of X for w zero, one obtains from (14·1) and (14·4)

$$\frac{\overline{X}}{X^*} = \frac{4}{3+e^{2w/kT}} \quad \text{(2nd approx.)}, \tag{14·5}$$

in place of the previous approximation of § 8

$$\frac{\overline{X}}{X^*} = \frac{2}{1+e^{w/kT}} \quad \text{(1st approx.)}. \tag{14·6}$$

The discrepancy between (14·5) and (14·6) will increase with w/kT and will be greatest (for $N_A = N_B$) at the critical temperature, which according to the first approximation is given by

$$e^{w/kT_c} = \frac{z}{z-2} = \frac{12}{10} \tag{14·7}$$

for a close-packed three-dimensional lattice.

Substituting from (14·7) into (14·6) and (14·5), one obtains

$$\frac{\mathrm{X}}{X^*} = \frac{2}{2{\cdot}20} = \frac{1}{1{\cdot}10} \quad \text{(1st approx.)}, \tag{14·8}$$

$$\frac{X}{X^*} = \frac{4}{4{\cdot}44} = \frac{1}{1{\cdot}11} \quad \text{(2nd approx.)}. \tag{14·9}$$

It is thus seen that in this particularly unfavourable case the two approximations differ by less than 1 % in the value obtained for $\bar{X}/X^*$.

15. *Zeroth approximation.* It has already been mentioned that the quasi-chemical equations for the X_{ij}'s can be solved explicitly only for binary mixtures. It is therefore expedient to derive approximate formulae for mixtures of more than two components. It is actually easy to obtain formulae in which the inaccuracy is of the order $(w_{ij}/kT)^2$. From the form of the quasi-chemical equations for the X_{ij}'s and from the definition of the X^*_{ij}'s, it is clear that all the $\bar{X}_{ij}-X^*_{ij}$ vanish when all the w_{ij}'s vanish. A formula can therefore be obtained for the free energy F of the required degree of accuracy by expressing $F(N_i, \bar{X}_{ij})$ as a power series in $(\bar{X}_{ij}-X^*_{ij})/X^*_{ij}$ and neglecting terms of order higher than the first. The Taylor series for $F(N_i, \bar{X}_{ij})$ is

$$F(N_i, \bar{X}_{ij}) = F(N_i, X^*_{ij}) + \Sigma_{ij} \frac{\partial F(N_i, X^*_{ij})}{\partial X^*_{ij}} \times (\bar{X}_{ij}-X^*_{ij}) + \text{terms in } \Sigma_{ij}(\bar{X}_{ij}-X^*_{ij})^2. \tag{15·1}$$

Since $-F(N_i, \bar{X}_{ij})$ is maximized with respect to $\bar{X}_{ij}$, it is known that $\partial F(N_i, X^*_{ij})/\partial X^*_{ij}$ is small of the first order in $(\bar{X}_{ij}-X^*_{ij})/X^*_{ij}$ and the second term on the right of (15·1) is small of the second order in $(\bar{X}_{ij}-X^*_{ij})/X^*_{ij}$. Hence to the required degree of accuracy

$$F(N_i, \bar{X}_{ij}) = F(N_i, X^*_{ij}). \tag{15·2}$$

Substituting from (6·1) and using (5·1) and (4·1), one obtains

$$\begin{aligned} F &= -kT\Sigma_i N_i \log f_i - kT \log g(N_i, X^*_{ij}) + z\Sigma'_{ij} X^*_{ij} w_{ij} \\ &= -kT\Sigma_i N_i \log f_i - kT \log g(N_i) + z\Sigma'_{ij} X^*_{ij} w_{ij} \\ &= -kT\Sigma_i N_i \log f_i \rho_i + kT\Sigma_i \log N_i! - \tfrac{1}{2}zkT \log (\Sigma_i q_i N_i)! \\ &\quad + (\tfrac{1}{2}z-1)\,kT \log (\Sigma_i r_i N_i)! + \frac{z\Sigma'_{ij} q_i N_i q_j N_j w_{ij}}{\Sigma_i q_i N_i}. \end{aligned} \tag{15·3}$$

The corresponding formula for the energy of mixing E is

$$E = \frac{z\Sigma'_{ij} q_i N_i q_j N_j w_{ij}}{\Sigma_i q_i N_i}. \tag{15·4}$$

By differentiating (15·3) with respect to N_i one obtains for the partial potentials μ_i and the absolute activities λ_i

$$\mu_i = kT\log\lambda_i = kT\log\frac{N_i}{f_i\rho_i\Sigma_j r_j N_j} + \tfrac{1}{2}zq_i kT\log\frac{\Sigma_j r_j N_j}{\Sigma_j q_j N_j} + zq_i\left\{\frac{\Sigma'_j q_j N_j w_{ij}}{\Sigma_j q_j N_j} - \frac{\Sigma'_{jk} q_j N_j q_k N_k w_{jk}}{(\Sigma_j q_j N_j)^2}\right\}. \quad (15{\cdot}5)$$

In the case of a binary mixture of A and B molecules (15·6) reduces to

$$\frac{\lambda_A}{\lambda_{A,0}} = \frac{r_A N_A}{r_A N_A + r_B N_B}\left\{\frac{N_A + r_B N_B/r_A}{N_A + q_B N_B/q_A}\right\}^{\frac{1}{2}zq_A}\exp\left\{\frac{zq_A w}{kT}\left(\frac{q_B N_B}{q_A N_A + q_B N_B}\right)^2\right\}, \quad (15{\cdot}6)$$

correct to the first order in w/kT.

As the treatment in §§ 1–13 is equivalent to Bethe's first approximation, I have referred to it as the first approximation, and I have referred to the more accurate treatment in § 14 as the second approximation. The formula of the present section must therefore for the sake of consistency be called the zeroth approximation. This is in accordance with the terminology used for regular solutions by Fowler & Guggenheim 1939, § 819). The zeroth approximation corresponds to Hildebrand's approximation for regular solutions and to Bragg and Williams's approximation in the treatment of order-disorder in alloys (Fowler & Guggenheim 1939, §§ 819, 1318).

16. *Formulae for z infinite.* The maximum value of z, corresponding to closest packing, is 12, and greater values of z have no physical meaning. It is, nevertheless, of interest to consider how the formulae for a binary mixture behave when $z \to \infty$ and $w \to 0$, keeping $zw = u$ finite, because the free energy takes a strikingly simple form, previously proposed by Flory (1942) in the special case $r_B = 1$.

Some care is needed in proceeding to the limit, because, for example, although $r-q$ tends to zero, $\frac{1}{2}z(r-q)$ remains equal to $r-1$. It is convenient to define 'volume fractions' x_A, x_B by

$$x_A = \frac{r_A N_A}{r_A N_A + r_B N_B}, \quad x_B = \frac{r_B N_B}{r_A N_A + r_B N_B}. \quad (16{\cdot}1)$$

After proceeding to the limit $z \to \infty$, it is seen that $x \to x_A$ and $1-x \to x_B$, but care must be taken not to replace x by x_A indiscriminately in terms containing z.

The formulae eventually obtained are

$$\log\frac{p_A}{p_{A,0}} = \log\frac{\lambda_A}{\lambda_{A,0}} = \log x_A + \frac{(r_B - r_A)\,x_B}{r_B} + x_B^2\frac{r_A u}{kT}, \quad (16{\cdot}2)$$

$$F - N_A kT\log\lambda_{A,0} - N_B kT\log\lambda_{B,0} = N_A kT\log x_A + N_B kT\log x_B + (r_A N_A + r_B N_B)\,x_A x_B u, \quad (16{\cdot}3)$$

$$E = (r_A N_A + r_B N_B)\,x_A x_B u. \quad (16{\cdot}4)$$

Differentiating (16·2) with respect to N_B, one obtains

$$\frac{\partial\log p_A}{\partial N_B} = -\frac{r_A^2 N_A + r_B^2 N_B}{(r_A N_A + r_B N_B)^2} + \frac{r_A N_A r_B N_B}{(r_A N_A + r_B N_B)^3}\,\frac{2r_A r_B u}{kT}, \quad (16{\cdot}5)$$

so that the equation

$$\frac{\partial \log p_A}{\partial N_B} = 0 \tag{16·6}$$

is equivalent to

$$r_A(r_A N_A)^2 - \left\{\frac{2r_A r_B u}{kT} - (r_A + r_B)\right\} r_A N_A r_B N_B + r_B(r_B N_B)^2 = 0. \tag{16·7}$$

The critical conditions obtain when (16·7) has equal roots in $r_A N_A / r_B N_B$. Thus

$$\frac{2u}{kT_c} = \left\{\frac{1}{\sqrt{r_A}} + \frac{1}{\sqrt{r_B}}\right\}^2, \tag{16·8}$$

$$\frac{x_A}{x_B} = \frac{r_A N_A}{r_B N_B} = \sqrt{\frac{r_B}{r_A}}. \tag{16·9}$$

In the particular case $r_B = 1$, the above formulae reduce to those of Flory (1942). They are less accurate than the zeroth approximation of § 15.

I am greatly indebted to Sir Ralph Fowler, F.R.S., for his interest, encouragement and constructive criticism during the preparation of this paper.

REFERENCES

Chang 1939 *Proc. Camb. Phil. Soc.* **35**, 265.
Flory 1942 *J. Chem. Phys.* **10**, 51.
Fowler & Guggenheim 1939 *Statistical Thermodynamics*. Cambridge University Press.
Fowler & Guggenheim 1940 *Proc. Roy. Soc.* A, **174**, 189.
Guggenheim 1935 *Proc. Roy. Soc.* A, **148**, 304.
Guggenheim 1938 *Proc. Roy. Soc.* A, **169**, 134.
Guggenheim 1944*a* *Nature, Lond.*, **153**, 255.
Guggenheim 1944*b* *Proc. Roy. Soc.* A, **183**, 203.
Orr 1944 *Trans. Faraday Soc.* **11**, 320.
Rushbrooke 1938 *Proc. Roy. Soc.* A, **166**, 296.

The initiation of long electrical discharges

By C. E. R. Bruce, M.A., B.Sc., A.M.I.E.E.

(Communicated by Sir Edward Appleton, F.R.S.—Received 15 *October* 1943—
Revised 11 *April* 1944)

The glow-arc transition theory of the first leader stroke of a lightning flash is further considered, and shown to be in accord with all the observations available, whereas the recombination theory is not considered to be. The greatly reduced potentials already indicated to be adequate to cause flashes, of the order of only 1 % of those required in a homogeneous field, and the corona currents from the leader stroke channel, are confirmed by a discussion of the records of the currents in, and electrostatic field changes caused by, the first leader, which further indicates that formulae derived from laboratory studies of corona currents from wires and from pointed conductors can be extrapolated to lightning voltages, i.e. over a range of about 1000 to 1.

1. INTRODUCTION

It was pointed out by Larmor (1914) that when electrical breakdown of a gas has once been started in a non-uniform electric field, the resulting discharge will extend beyond the region throughout which the critical electric field strength required for breakdown was initially exceeded. This effect was discussed by Wilson (1920) and Simpson (1929), the former of whom considered that the potential difference required to cause a vertical lightning flash between a thundercloud and earth might be thereby reduced by a factor of not more than six below that which would be required if the field had been homogeneous. Though the value obtained in this way for the potential of a thundercloud is that usually quoted (see e.g. Schonland 1939; Loeb 1939*a*; Allibone 1941), there are grounds for believing that some workers in the subject doubted whether the average gradients between cloud and earth to which such a difference of potential would give rise do in fact exist. Some degree of precision was recently given to these latter ideas (Bruce & Golde 1941).

The first indication that the pre-discharge fields are comparatively small was obtained when the charge distributed along the leader stroke channel was considered in relation to the latter's distributed capacitance. This charge distribution was derived from an interpretation of oscillographic records of the current in a lightning flash, in the light of observations of the velocity of the return stroke, the latter having been obtained by Schonland and his collaborators (1934, 1935, 1938), from their photographs of lightning flashes taken with a Boys camera. The duration of the initial high-current portion of the lightning current wave is about 100 μsec., agreeing with the time required for the return stroke to reach the cloud charge, the charge distributed along the leader stroke channel having been neutralized during its passage. This latter charge is found by integration of the oscillograms to be about 1 coulomb, while the capacitance of such a vertical conductor just not in contact with the earth and 2·5 km. long, from a formula given by Goodlet (1937), is about $1{\cdot}28 \times 10^4$ $\mu\mu$F. Its potential is therefore $7{\cdot}8 \times 10^7$ V, giving an average field between cloud and earth along the actual path taken by the lightning flash, of 312 V/cm., while the shape of the current wave and the distribution of capacitance along the conductor together indicate that the field is fairly uniformly distributed.

Again, the results obtained by Schonland and his collaborators indicate that there are on an average about three successive strokes in a flash, while the average time interval between these strokes is about $\frac{6}{100}$ sec. Since the velocity of propagation of the leader stroke process is about 3×10^7 cm./sec., this indicates that the thundercloud charge is spread over an area of the order of 100 sq. km. This latter charge, the total charge brought to earth in a flash, is on an average about 30 coulombs, so that the average field is thus about 200 V/cm., and the potential of the thundercloud charge about $4{\cdot}5 \times 10^7$ V.

A similar conclusion as to the dispersion of the charge in the thundercloud is arrived at from a consideration of the duration of thunder. The data given by Wilson (1920) and Mathias (1924) show that in the case of near flashes this lasts for over 30 sec., and in some cases for over 40 sec., indicating that the diameter of the charged area is of the order of more than 10 km., in agreement with the above estimates.

The average pre-discharge field between cloud and earth is thus apparently only of the order of one or a few hundred volts per centimetre, or only about 1 % of that required to cause breakdown in a uniform field. This estimate is, however, of the same order as that of the field changes and pre-discharge fields actually recorded at the earth (Wilson 1920; Wormell 1939; Simpson & Scrase 1937; Appleton & Chapman 1937), and as that of the fields recorded during the ascent of test balloons through thunderclouds, when lightning flashes occurred nearby (Simpson & Scrase 1937; Simpson & Robinson 1941). Through the courtesy of Sir George Simpson, the author was able to examine the records obtained in the last-mentioned investigations, to see if there was any evidence for a gradual increase of the electric field with increase of height above the ground, but none was found.

It has generally been considered that the individual strokes of a flash represent the successive discharge to earth of separate volumes of charge in the cloud (Schonland 1938). This complication of the simpler distribution suggested above, would, however, appear to be unnecessary, and is rendered improbable by the occasional occurrence of flashes comprising large numbers of separate strokes, as many as 27 (Schonland, Malan & Collens 1935), and even 40 (Larsen 1905) and 42 (Matthias 1929) having been observed in one flash. A likelier explanation would appear to be that each stroke results from the meeting of a leader progressing from the initial discharge channel into the cloud, with one which has originated at a local elongation or protuberance from the main volume of charge, similar to that responsible for the initiation of the first stroke to earth. The charge deposited along this second leader will then be rapidly distributed along the channel to earth, giving the phenomenon of the continuous dart leader of subsequent strokes (Schonland *et al.* 1935), and thereafter neutralized during the ensuing return stroke. In the absence of such local field concentrations, the channel or channels from the point of origin of the flash will simply progress through the cloud neutralizing the charge therein by a gradual process (corona discharge), and it may be noted that the largest charge hitherto recorded oscillographically (McEachron 1939) was practically entirely conducted to earth in just such a relatively continuous manner.

So far as the field changes caused by successive strokes are concerned, the above interpretation will give rise to the same phenomena as would be observed were the total charge in fact initially segregated into separate pockets. Until further evidence is obtained, therefore, it would appear to be sufficient to consider the comparatively simple distribution of charge suggested above, and in what follows a mechanism is suggested to account for the propagation of an electrical discharge through a gas in which the field strength, apart from a very localized concentration at the point of origin of the discharge, is of the order of that suggested above, and in the process further precision is given to the values of the fields involved.

2. IMPORTANCE OF CORONA CURRENTS

An important aspect of these very long electrical discharges has recently been pointed out (Bruce 1941), namely, that heavy corona currents must flow laterally from the leader stroke channel during its growth towards the opposite electrode. Since, apart from the relatively small potential gradient required for the maintenance of arc conditions, the potential of the initiating electrode is being propagated into the intervening space, the potential difference between the streamer and the surrounding space must build up to many millions of volts, even with the much smaller field strengths now obtained. Such potentials are in fact observed when the streamer makes contact with an overhead transmission line (Pittman & Torok 1931). It is shown herein (§ 6) that in the case of the lightning discharge these lateral corona currents must build up to the order of hundreds or even thousands of amperes, during the progression of the leader stroke. That such currents do, in fact, flow has been confirmed by direct oscillographic records of the current in the upward leader strokes from the Empire State Building (McEachron 1939), and also indirectly by records of the electrostatic field changes caused by the leader stroke of nearby flashes (§§ 4, 6).

It would appear that the existence of currents of this order rules out the possibility of the ageing of the ionization in the leader stroke channel, envisaged by Schonland (1938), and also the recombination theory of the stepped structure of the first leader stroke as it has been developed by Meek (1939) and Loeb (1939*b*).

It is also evident that the only type of electric discharge capable of carrying the large currents which are thus seen to flow steadily in the leader stroke is the arc discharge, and the realization of this fact has led to a new theory of the stepped leader stroke.

3. MECHANISM OF THE STEPPED LEADER STROKE

The streamers which arise, for example, from the top of the Empire State Building, prior to the occurrence of a lightning flash, must start as brush or glow discharges carrying minute currents, so that at some stage in their development when an upward stroke occurs from the Building, the sudden transition to arc conditions must take place. Again, if the fresh start of the pilot streamer (Schonland 1938) after the occurrence of a dart is considered, the same sequence must ensue. The current required to maintain the lateral corona must increase as the pilot streamer lengthens, so that an increasing current must flow through the tip of the preceding dart, or initiating point, until a total current of the order of an ampere is reached, at which point the sudden transition from glow to arc conductivity is known to occur (Todd & Browne 1930; Attwood, Dow & Krausnick 1931; Fan 1939). This condition is represented diagrammatically in figure 1, in which I_T is this critical transition current. The reason for the complete breakdown of the whole additional length of the streamer, *AB*, to arc conditions at this point in its history, is that a position of instability has been reached, since the currents flowing along the streamer from *A* to some point *A′*, near *B*, are already large enough to maintain arc conditions, with a voltage gradient only about one-tenth of that required for the conduction of the same current in the glow regime, if only arc conditions could be initiated therein. This follows from the observation (Attwood *et al.* 1931) that, when once established, arc conductivity is maintained down to currents of the order of only one-tenth of those required to cause glow to arc transition, before transition back from arc to glow occurs. The disturbance and redistribution of space charge caused by the initiation of arc conditions at *A*, is sufficient to trigger the change-over to the new conditions of stability throughout the newly formed channel, though the change will not be permanent beyond *A′*. It is this 'hysteresis' effect, so to speak, which accounts for the observed dart structure.

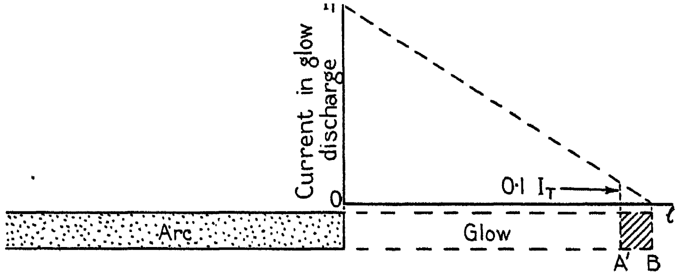


FIGURE 1. Diagram illustrating formation of leader stroke darts. (I_T is the current required for glow to arc transition.)

Up to the instant of transition the characteristics of the newly formed streamer will be at best those of the glow discharge, i.e. the voltage gradient required to maintain it will be of the order of several hundred volts per centimetre, whereas at the transition it will be suddenly reduced to the order of tens of volts per centimetre, and will have a pronouncedly negative volt-ampere characteristic, so that as the current through it rises with the increasing length of the channel, the gradient required to maintain the discharge current will be still further reduced. Another important feature of the new regime is that the streamer channel after the transition has the characteristics of an arc plasma, and is from most points of view simply an extension of the original conductor. It is this important characteristic, together with the greatly reduced potential gradient required for the maintenance of arc conductivity, which explains the 'self-propagating' nature of the discharge.

4. ELECTROSTATIC FIELD CHANGES CAUSED BY THE LEADER STROKE

There are several ways in which the new theories of the corona currents and the mechanism of the leader progression can be checked quantitatively against the observed results. Perhaps the most striking is the comparison of the calculated and observed electrostatic field changes caused by the leader stroke.

First, as regards the potential required to initiate such a discharge: The largest pre-discharge field which has hitherto been observed is about 100 V/cm. (Wormell 1939), i.e. well below the estimates given in § 1 above. The Empire State Building short-circuits, as it were, 1250 ft. of this field, so that if fields of this order existed around the Building, then the potential difference between its top and the surrounding space would be about 4,000,000 V. However, it will be shown in § 6 that flashes are apparently initiated by the Building in fields which would not normally result in lightning flashes from the cloud, the pre-discharge field being in the case of these upward flashes apparently only about 50 V/cm. With such fields the potential difference between the top of the Building and the surrounding space is only about 2,000,000 V, so that this would appear to be the potential difference required to initiate a positive discharge in such a case, for so far only flashes in which the Building was the positive electrode have been observed. It will be shown in § 7 that a similar value for this critical potential difference is indicated by extrapolation from laboratory data on the magnitude of currents from pointed conductors, since the latter suggest that currents of the order of that required for the occurrence of glow to arc transition will flow from the point at these potentials. Though extrapolation over such a wide range of potential may appear at first sight uncertain, to say the least of it, nevertheless the agreement of theory and observation is rendered less surprising by what follows on the magnitudes of the lateral corona currents, which are likewise found to agree with the values obtained by extrapolation from laboratory data.

Once the streamer is started, its potential, V_l, relative to the surrounding space builds up at the rate of X_1 V/cm. travelled, if the small voltage gradient required to maintain the arc, X_a, is neglected, where X_1 is the value of the pre-discharge electric field between cloud and ground. At a point distant l from the start, it is therefore given by

$$V_l = V_0 + l(X_1 - X_a) \doteqdot V_0 + lX_1,$$

where V_0 is the initial potential difference between the point at which the streamer originates, and the surrounding space.

The corona current per centimetre length of the discharge channel at this point will be

$$I_c = k(V_0 + lX_1)^2, \tag{1}$$

assuming that this current is proportional to the square of the conductor voltage (Whitehead 1927), the potential which marks the onset of corona being in this case relatively negligible. The rate of change of electric moment at time t is therefore

$$\frac{d\mu}{dt} = \int_0^{vt} k(V_0 + lX_1)^2\, 2l\,dl, \tag{2}$$

where v is the velocity of propagation of the leader stroke, and the total electric moment destroyed in time t is

$$\mu = \int_0^t \int_0^{vt} 2kl(V_0 + X_1 l)^2\, dl\, dt. \tag{3}$$

The electrostatic field change at distance r from the flash due to this change of electric moment is given by

$$\epsilon_s = \frac{\mu}{r^3}, \tag{4}$$

in which μ and ϵ_s are in electrostatic units, so that

$$\epsilon_s = \frac{9 \times 10^{13}}{r^3} \int_0^t \int_0^{vt} 2kl(V_0 + X_1 l)^2\, dl\, dt \text{ V/m.}, \tag{5}$$

l and r being in cm., v in cm./sec., V_0 and X_1 in V and V/cm. respectively, and ϵ_s in V/m.

No allowance has been made in the above for branching of the leader stroke, or for the fact that the general field through which the leader is advancing decreases somewhat as charge is lowered from the cloud by the leader stroke. As regards the former, more than half the total number of flashes appear to be unbranched (Schonland & Collens 1934; Schonland *et al.* 1935), and for these the present analysis holds. The average number of branches in branched flashes is about 3 (Schonland *et al.* 1935), but not all these are in process of formation at the same time. It is therefore unlikely that allowance for branching would materially alter the results deduced herein. The effect would be to reduce still further the values obtained for the electric field.

The allowance to be made for the change of electric field strength during the leader stroke will also be small. This field change has been measured by Appleton & Chapman (1937) and by Schonland, Hodges & Collens (1938) who found it to be 33 and 25 %, respectively, of the total field change due to the stroke. As there are on an average three strokes in a flash, the field change caused by the first leader stroke is thus only of the order of 10 or 20 % of the total change caused by the flash. Again, the increase of the estimated initial potential required to compensate for the decrease of the field during the progression of the leader stroke will be further reduced by the fact that the effects considered increase as the square of the potential difference between the channel and the surrounding space.

Thus allowance for these two factors would certainly not affect the order of the values deduced herein for the potential of the thundercloud charge, the effect of each being small and its direction opposite to that of the other.

Oscillographic records of ϵ_s for near strokes have been obtained by Appleton & Chapman (1937), so that the above is an equation for k when V_0, X_1 and the height of the thundercloud charge are known. This last varies somewhat, but may be taken to be nearly 2·5 km. (Simpson & Scrase 1937). The value of ϵ_s at the end of the leader stroke, i.e. immediately prior to the sudden change due to the return stroke, was measured in the two oscillograms given by Appleton & Chapman (their figure 6, (i) and (ii)), and the values of k calculated in each case, corresponding to values of X_1 of the order of a few hundred volts per centimetre. The resulting curves are shown in figure 2, from which k will be seen to be of the order of 10^{-16} or 10^{-17}.

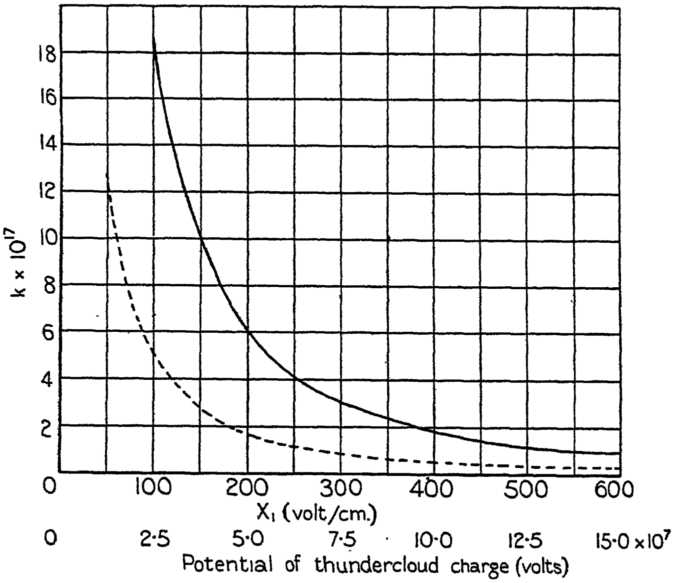


FIGURE 2. Variation of corona constant k with potential of thundercloud charge. Curves based on figure 6 (i) and (ii) of Appleton & Chapman (1937). Full line, figure 6 (i); broken line, figure 6 (ii).

That these two records represent fairly normal conditions is seen from a comparison of the fields and durations involved with the average conditions found by these and other investigators. The flashes in question both occurred at about 10 km. distance, and the total field change caused by the stroke was about 1000 V/cm. in each case. The average value obtained by Appleton & Chapman at this distance was about 650 V/cm., while the average field change caused by a complete flash at the same distance was found by Wormell (1939) to be about 1100 V/cm., thus indicating that if anything the field changes in the two oscillograms in question are on the high side, but probably do not exceed the average by a factor of more than about 2.

Again, as to the durations of the field changes, these indicate for the velocity of the leader stroke values of about $2{\cdot}3 \times 10^7$ and less than 7×10^7 cm./sec. respectively in the two cases, as compared with Schonland's mean value of about $3{\cdot}8 \times 10^7$ cm./sec., and McEachron's mean value of about $2{\cdot}5 \times 10^7$ cm./sec.

Several investigations of the corona currents from wires have been made and from them the value of k can be determined. Thus Watson's data on d.c. corona (1910) yield a value of $k = 4{\cdot}2 \times 10^{-17}$, while Peek's data (1915) give values of k somewhat higher than this.

This agreement between the calculated and measured values of k would appear both to confirm the existence of these corona currents and the greatly reduced values of the thundercloud potential now obtained.

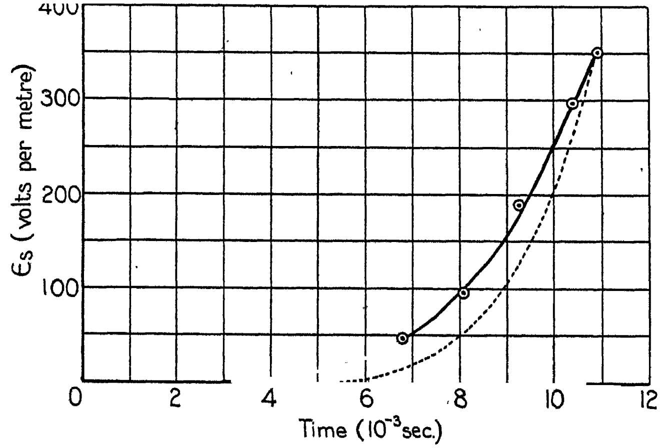


FIGURE 3. Electrostatic field change due to leader stroke. Observed (Appleton & Chapman, 1937, figure 6 (ii)) and calculated. Points measured from oscillogram. Broken line, original calculation; full line, calculated with new time base.

When the variation of ϵ_s with time was studied, there appeared to be a divergence between the calculated and observed values. (Compare the full and broken line curves of figure 3.) However, on further consideration it appeared that since, on the present view, a lightning flash results from a breakdown originating at a field concentration in a field which is elsewhere fairly uniform, the records obtained of the change of electrostatic field with time should be geometrically very similar, provided that there is sufficient charge available to maintain the discharge. As it has just been seen that the field changes caused by the two strokes in question are both above the average, the latter condition would appear to be satisfied. Comparison of the aforementioned records (Appleton & Chapman 1937, figure 6 (i) and (ii)) showed that such a similarity did not exist between these two records, and that the initial quiescent period in their figure 6 (ii), before the trace leaves the zero line, already of long duration, should be longer still. This at once suggested that the oscillograph had in this case only tripped some time after the lightning flash had actually started, owing to the weakness of the initial pulses. A lower limit to this time lag was obtained from a comparison of the two records by making the assumption that no lag had occurred in figure 6 (i). The required increase in the time base of figure 6 (ii) was thus found to be about 3·4 msec., the total duration of the leader stroke being then 10·8 msec. Using this new time base, complete agreement is obtained between the calculated and observed field changes, as is seen in figure 3, in which the circles represent the values measured from the oscillogram, while the full line curve is that calculated from equation (5) above using the new time base. The theory thus offers a satisfactory explanation of these records, which had remained unaccounted for since they were first obtained by Appleton & Chapman.

5. Determination of the thundercloud potential

It will be seen that on the present theory there is no well-defined potential required for the occurrence of a lightning flash, since the higher the average field the less will be the discontinuity necessary to initiate the 'self-propagating' discharge. Thus upward flashes will be initiated by high buildings in fields which are approximately inversely proportional to the heights of the buildings, and the same will apply to elongations of space charge in the cloud. Indeed, there would appear to be no reason why, in the extreme case, the average field need be much greater than that required for the maintenance of an arc, i.e. of the order of 10 V/cm. in air at atmospheric pressure, and it will be shown in § 6 that, in the case of flashes to the Empire State Building, the average field between cloud and ground is probably only of the order of 50 V/cm.

The results of the previous section afford a method of determining the potential of the thundercloud charge in individual cases from records of the electrostatic field change during the leader stroke, if at the same time the height of the charge is determined by the method of sounding balloons adopted by Simpson. From each of the former records, together with the known height of the charge, a curve corresponding to those in figure 2 can be derived, giving k, the corona constant, in relation to X_1. This k must be a constant characteristic of these long-discharge channels, the value of which will become apparent when more records are available for study. When the value of k has thus been determined, the curve for each flash will then yield a value for the average gradient below the thundercloud charge for the flash in question. As an illustration, had the heights of the charge in the two cases dealt with in figure 2 been found each to have the average value of 2·5 km., then if we provisionally adopt the value of k found experimentally, viz. about 4×10^{-17}, the cloud potentials would have been $6{\cdot}3 \times 10^7$ and $3{\cdot}0 \times 10^7$ V respectively. From the combination of the results of the two investigations suggested above it will likewise be possible to determine the velocity of the leader stroke, and hence the variation of the latter with the initial field strength. The indications of the comparison of the above two records are that velocity decreases with decrease in field strength, and in this connexion it is of interest to note that the long durations observed by McEachron

(1939) in the case of flashes to the Empire State Building may be due to this cause, the flashes having been initiated by the Building in fields which are smaller than those required to cause normal flashes from cloud to earth.

A further interesting deduction from the present theory of the initiation and propagation of a lightning flash is that the protective range of a lightning conductor or tall building will not be constant for all lightning flashes, but will increase with the potential difference between the leader stroke channel as it approaches the earth and the surrounding space. For different leader strokes along any one vertical path, the induced potential of the space immediately surrounding the lightning conductor will reach the value necessary for the initiation of a continuing streamer when the leader stroke is at different heights, so that the earlier this occurs the more chance is there of this streamer diverting the downcoming leader towards the lightning conductor.

6. CURRENT IN THE LEADER STROKE

A further check of the present theory is derived from a consideration of the magnitude of the current in the leader stroke, since this has been measured oscillographically in strokes to the Empire State Building by McEachron (1939). From equation (1) it follows that the current required to maintain the corona discharge from the channel at time t from the start is given by

$$I_t = \int_0^{vt} k(V_0 + X_1 l)^2 \, dl \tag{6}$$

$$= kvt(V_0^2 + V_0 X_1 vt + 1/3 X_1^2 v^2 t^2), \tag{7}$$

which, for the flashes considered, rises to values of the order of 5000 amp. These values appear high, especially in view of the fact that previous estimates of the steady currents during the leader stroke were of the order of 0·1 amp. to several amperes (Meek 1939; Schonland 1938). However, currents of the order of those now obtained are essential to account for the field changes observed by Appleton & Chapman, whose results are consistent with those of Wilson (1920), Wormell (1939) and Laby and others (1940) as regards the magnitudes of the field changes observed, and with those of Schonland and others (1938) as regards the fraction of the total field change which occurs during the leader stroke.

The values obtained for the currents in the leader stroke are certainly considerably higher than those observed oscillographically by McEachron (1939) in upward strokes from the Empire State Building, which last rose only to about 500 amp. This is not surprising, however, since flashes must have occurred from the Building at field strengths which would not ordinarily have caused flashes to occur from the cloud, so that a lower value of the current is to be expected. That this in fact happened is seen from the very high number of flashes, as many as 11 or 19, which struck the Building in *one day*, numbers very much greater than the number of flashes which would have occurred over an area of normal country equal to that 'protected', in an electrical sense, by the Building. It has been suggested (Meek

1941) that these oscillograms only indicate the peaks of the current pulses. It is to be noted, however, that the envelopes of the troughs themselves reach about 300 and 500 amp. respectively during the leader stroke in the two of McEachron's oscillograms to which reference is made by Meek.

The currents observed by McEachron in fact enable the average field required to cause a flash from the Building to be calculated, simply by regarding (7) as an equation to determine X_1. The average gradient is thus found to be about 55 V/cm., using the same values of the corona constant and height of the cloud charge as in the previous section, so that the total potential required is thus only of the order of $1{\cdot}5 \times 10^7$ V. This conclusion could be checked by recording the electrostatic fields and field changes during thunderstorms in the vicinity of the Building, either on one of the surrounding buildings, using the method of Appleton & Chapman, or in the space below the thundercloud, by the use of sounding balloons.

This field, together with the height of the Building, yields for the potential difference between the top of the Building and the surrounding space necessary to initiate the discharge in this case a value of the order of 2,000,000 V. So far only flashes in which the Building was the positive electrode have been obtained, so that the above result applies at present for this polarity only, and for a field concentration equal to that caused by the Building.

It will be noted that no account has been taken of the fact that, whereas all the downward leaders so far investigated have been negative, all the upward leaders from the Empire State Building have been positive. Though positive corona currents from wires exceed negative, there are so far no results to indicate whether this applies also in the case of these discharge channels. The fact that positive lightning currents tend to be greater than negative (Bruce & Golde 1941) indicates, as Allibone (1941) pointed out recently, that the same applies in this case, and that the higher positive lightning currents may result from the neutralization during the return stroke of the greater charge deposited around a positive channel by the larger corona currents. As, however, the velocity of positive return strokes has not yet been measured, there remains the possibility that it is higher than that of negative strokes, which would equally well explain the higher positive currents. It is to be hoped that photographs of positive flashes will ultimately be obtained by Schonland and his co-workers and settle this point. Meanwhile, it appears quite probable that the difference between the corona currents from positive and negative wires is due to the difficulty of getting free electrons out of the metal conductor in the latter case, a difficulty which will not arise when the conductor is a gas discharge channel, so that it is possible that the pronounced polarity effect will not be found in the latter case.

The general agreement between the theoretical and observed values of the current in the leader stroke—the former tending to be high if anything—affords further evidence in favour of the low value of the thundercloud potential, in view of the dependence of the corona currents on the square of the potential. Since the values of the latter hitherto accepted were about 100 times greater than those now deduced, the leader stroke currents would be 10,000 times those actually observed.

7. CURRENT FROM A POINTED CONDUCTOR

It is of interest to compare the result obtained in the last section on the initial potential of the space surrounding the top of the Empire State Building at the start of a flash, with results obtained in investigations of the currents from pointed conductors, since this current should be, on the theory now put forward, that required to yield a brush discharge current sufficient to cause the transition from glow to arc conditions, i.e. a current of about 0·5 or 1 amp. Both Warburg (1899) and Zeleny (1907, 1908, 1941) have found that the current from a point can be represented by an expression of the form

$$I = cV(V - V_i), \tag{8}$$

V_i being the potential at which the discharge starts, and c a constant. In the present case this becomes simply

$$I = cV_0^2. \tag{9}$$

Since for $V = V_0 = 2 \times 10^6$ V, I should be equal to approximately 1 amp., c in this case is equal to $2{\cdot}5 \times 10^{-13}$, a value agreeing very well with those found by the abovementioned investigators, Warburg's values being $1{\cdot}48 \times 10^{-13}$ and $4{\cdot}83 \times 10^{-13}$, while Zeleny's values lie between $2{\cdot}6 \times 10^{-13}$ and $4{\cdot}2 \times 10^{-13}$. It would thus appear on the present data that both the corona constant and this constant on which the discharge from a point depends, can be extrapolated through a range of about 1000 to 1 in potential.

8. LENGTH OF THE FIRST LEADER STROKE DARTS

It was seen in § 6 that the potential difference between the top of the Empire State Building and the surrounding space at the start of a lightning flash from the Building is probably about 2,000,000 V, so that the lateral corona current from each centimetre length of the discharge channel, kV_i^2, is initially about $1{\cdot}6 \times 10^{-4}$ amp. For this current to build up to 0·5 or 1 amp. a length of 3000–6000 cm., or 100–200 ft., is required. By the time the discharge reaches the bottom of the cloud, which may be taken as at about 1·25 km. (Simpson & Scrase 1937), V_i will have increased to about 5·6 million volts,* so that the length of the darts will be reduced to about 400–800 cm., or 13–26 ft. We should thus expect the dart lengths observed by McEachron to lie between about 15 and 150 ft. The limits quoted by him are 19 and 60 ft. in quite good agreement with the theoretical values.

The comparison cannot be made precisely in the case of Schonland's observations since data on the initial fields are lacking, but indicates that the latter must be less than 100 V/cm.

9. REILLUMINATION OF THE DARTS

As has already been mentioned, once arc conditions have been initiated in a discharge channel owing to the increase of the current therein beyond a certain value, they are apparently maintained when the current is reduced down to a value of the order of one-tenth of that required to cause glow to arc transition in the first place. Again, the current which flows at any point in the leader stroke channel is that required to maintain the corona currents from the channel beyond the point in question. It follows from these two considerations that, after transition to arc conditions has occurred throughout the whole of the newly formed length of pilot streamer, the currents flowing in the last tenth of this length will be insufficient to maintain arc conditions therein, and retransition back to glow conductivity will occur. Consequently, when the new step is formed, about one-tenth of the old step will be reilluminated at the same time. It is therefore interesting to note that early in their investigation Schonland and his co-workers (1935) gave the following as the second of nine characteristics of the darts: 'A bright step generally starts a little way back on part of the track formed by the previous step, so that the portion of the step which is entirely new is about 90 % of the whole...', exactly as the theory leads us to expect.

---

* The potential of the channel relative to the surrounding space will build up at the rate of $X_1 - X_a = 55 - 10 = 45$ V/cm.

10. THE LONG SPARK

The long laboratory spark will, it is hoped, be dealt with on the present lines in the near future, but several of its outstanding characteristics may be mentioned as being in accord with the theory now advanced. In the first place, the currents in its leader stroke are of the order of 0·5–1 amp. (Tamm 1928; Allibone & Meek 1938; Komelkov 1940), as required if transition from glow to arc conditions were the criterion governing its initiation and progression.

Again, in some of Allibone and Meek's experiments, not only has the gaseous conductor gone over to the arc regime, but the conditions required for arc conductivity have in addition been established at the electrode surface, which is seen to be glowing in their figures 1 (*d*), 8 (*a*), and in others of their photographs, during the leader stroke. The corona discharge from the leader stroke channel is likewise apparent, and is repeatedly referred to by the authors as a 'shower of discharge'.

11. TYPES α AND β LEADER STROKES

Schonland and his co-workers (1938) have found that in about one-third of the lightning flashes which they photographed the velocity of the leader stroke decreased during the second half of its travel from cloud to earth, and have attributed the higher values just below the cloud to the existence there of a volume of space charge.

The present consideration of the corona currents from the discharge channel suggests, however, that the cause of the slowing down may be the inability of the discharge processes in the thundercloud to supply the charge required to maintain these rapidly increasing currents. Indeed many flashes never reach the earth at all, indicating that the cloud charge may actually be exhausted in supplying these currents.

12. THE RECOMBINATION THEORY OF THE STEPPED LEADER STROKE

Certain criticisms of the recombination theory of the stepped nature of the first leader stroke have already been offered (§ 2). It is, further, difficult to understand why the conductivity in the channel should decay, when once the latter has been formed, if, as the theory assumes (Meek 1939), a field of the order of that required for breakdown is available to restart the conduction processes after this assumed decay. It might perhaps be argued that the charge in the neighbourhood of the channel in the cloud becomes exhausted, but this explanation would fail in the case of the upward leaders from the Empire State Building, in which the steps occur in exactly the same manner as they do in the downward leaders from the cloud.

The theory would also appear to leave the bright ends of the darts unexplained, as well as the reillumination of the last tenth of the preceding dart. There is also the important consideration that in no case has the return stroke been observed to find any difficulty in progressing along the whole leader stroke channel at a very high velocity. It would be a strange coincidence if no case had been photographed in which the leader stroke met the earth when recombination had progressed far towards the exhaustion of conductivity.

13. CONCLUSIONS

(*a*) The low values of the electric field required to cause a lightning flash, of the order of one or a few hundred volts per centimetre, put forward in an earlier paper, are confirmed.

(*b*) Corona currents from the discharge channel are shown to account for the currents observed in the leader strokes of lightning flashes from the Empire State Building, and for the records of the electrostatic field changes caused by the leader strokes of flashes to earth.

(*c*) The glow to arc transition theory of the darts in the first leader stroke of a lightning flash is shown to be in accord with the observed characteristics of the darts, whereas the recombination theory is not considered to be.

(*d*) Flashes probably occur to the Empire State Building when the average field between cloud and earth is only of the order of 50 V/cm.

(*e*) The laboratory values of the corona constant and the constant on which the discharge from a point depends would appear to allow of extrapolation through a range of 1000 to 1 in potential.

(*f*) The potential of the thundercloud charge can be determined from records of the electrostatic field change caused by the leader stroke, together with a measure of the height of the charge, and hence the variation of the leader stroke velocity with the initial potential gradient, and the energy in individual flashes can likewise be determined.

(*g*) The zone of protection afforded by a lightning conductor may be expected to increase with the current in a lightning flash.

References

Allibone, T. E. 1941 *Quart. J. R. Met. Soc.* **67**, 327.
Allibone, T. E. & Meek, J. M. 1938 *Proc. Roy. Soc.* A, **166**, 97.
Appleton, E. V. & Chapman, F. W. 1937 *Proc. Roy. Soc.* A, **158**, 1.
Attwood, S. S., Dow, W. G. & Krausnick, W. 1931 *Trans. Amer. Inst. Elect. Engrs*, **50**, 854.
Bruce, C. E. R. 1941 *Nature, Lond.*, **147**, 805.
Bruce, C. E. R. & Golde, R. H. 1941 *J. Instn Elect. Engrs*, **88**, 487.
Fan, H. Y. 1939 *Phys. Rev.* **55**, 769.
Goodlet, B. L. 1937 *J. Instn Elect. Engrs*, **81**, 487.
Komelkov, W. S. 1940 *Elektrichestvo*, no. 9, 24.
Laby, T. H., McNeill, J. J., Nicholls, F. G. & Nickson, A. F. B. 1940 *Proc. Roy. Soc.* A, **174**, 145.
Larmor, J. & Larmor, J. S. B. 1914 *Proc. Roy. Soc.* A, **90**, 312.
Larsen, A. 1905 *Ann. Astrophys. Obs. Smithson. Instn*, p. 119.
Loeb, L. B. 1939 *Fundamental processes of electrical discharge in gases*: (*a*) p. 541, (*b*) p. 544. London: Chapman and Hall Ltd.
McEachron, K. B. 1939 *J. Franklin Inst.* **227**, 149.
Mathias, E. 1924 *Traité d'Électricité Atmosphérique et Tellurique*. Paris: Les Presses Universitaires de France.
Matthias, A. 1929 *Elektrotech. Z.* **50**, 1469.
Meek, J. M. 1939 *Phys. Rev.* **55**, 972.
Meek, J. M. 1941 *J. Instn Elect. Engrs*, **88**, 509.
Peek, F. W. 1915 *Dielectric phenomena in high voltage engineering*, p. 132. New York: McGraw Hill.
Pittman, R. R. & Torok, J. J. 1931 *Trans. Amer. Inst. Elect. Engrs*, **50**, 568.
Schonland, B. F. J. 1938 *Proc. Roy. Soc.* A, **164**, 132.
Schonland, B. F. J. 1939 *Terrestrial magnetism and electricity*, p. 663, ed. J. A. Fleming. New York and London.
Schonland, B. F. J. & Collens, H. 1934 *Proc. Roy. Soc.* A, **143**, 654.
Schonland, B. F. J., Hodges, D. B. & Collens, H. 1938 *Proc. Roy. Soc.* A, **166**, 56.
Schonland, B. F. J., Malan, D. J. & Collens, H. 1935 *Proc. Roy. Soc.* A, **152**, 615.
Simpson, G. C. 1929 *J. Instn Elect. Engrs*, **67**, 1269.
Simpson, G. C. & Scrase, F. J. 1937 *Proc. Roy. Soc.* A, **161**, 309.
Simpson, G. C. & Robinson, G. D. 1941 *Proc. Roy. Soc.* A, **177**, 281.
Tamm, R. 1928 *Arch. Elektrotech.* **19**, 235.
Todd, F. C. & Browne, T. E. 1930 *Phys. Rev.* **36**, 732.
Warburg, E. 1899 *Ann. Phys., Lpz.*, **67**, 72.
Watson, E. A. 1910 *Electrician*, **64**, 707.
Whitehead, S. 1927 *Dielectric phenomena: electrical discharges in gases*, p. 135. London: Ernest Benn Ltd.
Wilson, C. T. R. 1920 *Phil. Trans.* A, **220**, 73.
Wormell, T. W. 1939 *Phil. Trans.* A, **238**, 249.
Zeleny, J. 1907 *Phys. Rev.* **25**, 305.
Zeleny, J. 1908 *Phys. Rev.* **26**, 129.
Zeleny, J. 1941 *J. Franklin Inst.* **232**, 23.

Address of the President Sir Henry Dale, O.M., G.B.E., at the Anniversary Meeting, 30 November 1944

The annual number of *Obituary Notices of Fellows of the Royal Society* published today, and the names which have just been read to us, remind us of the losses the Society has suffered.

Allow me first to make brief mention of the last service rendered to the Society by one who had long been devoted to its interests, and whose name is among those of the Fellows whom death has taken from us during the past year. Sir Henry Lyons, who was our Foreign Secretary for a year, and then achieved so much for the Society in his full term of service as Treasurer, had acquired in that period a deep interest in the handling of the Society's business and in the changes in its structure and its administration over the centuries of its history. He devoted the last four years of his life, under conditions which must have deterred any less resolute enthusiast, to the writing of an historical account of the administration of the Royal Society under its Charters, and at the time of his death, last August, he was eagerly awaiting its publication, which war-time difficulties had long delayed. This long expected contribution to our history was published a few weeks ago, and the Society will welcome and cherish it, not only as a record of value and interest in itself, but in memory of one to whose devoted labours the Society and its Fellows owe so much.

Then let me say, on behalf of all the Fellows, a word of farewell and grateful acknowledgment to a member of our permanent staff who, after long and efficient service, is retiring from full duty. Mr Ronald Winckworth has for nearly 20 years, with only a short interruption, acted as assistant editor of our publications, and as the courteous helper and friend of all who have contributed to them. It may be doubted whether anyone else for so long a term has read, with a critical attention, every word and symbol of Philosophical Transactions and Proceedings, both A and B, and survived to accept the applause which such a feat deserves. Though our friend, Mr Winckworth, can no longer accept the full burden of that vigilant responsibility, you will be glad to know that we shall still be able to call upon his experience for less frequent consultation. It is a matter for satisfaction, too, to think that science will not be without compensation for its loss over the wide range of our publications, since Mr Winckworth will have a greater freedom to concentrate his interest on his own special study of the Molluscs, on which he is one of our leading experts. He may be assured that he takes with him the gratitude and good wishes of us all.

A year ago I reported to the Society that our Biological Secretary, Professor A. V. Hill, had left us on an important mission to India. The Council of the Society, at the invitation of the Indian Government, had nominated Professor Hill to visit India to see its problems for himself, so that he might offer his advice on scientific matters in general and, in particular, on the adoption for India of a new and progressive programme of research and enterprise in science and its applications. From all sources—from the Viceroy and the Secretary of State with their official colleagues and counsellors, and directly from our own colleagues, the Indian men of science—we have had evidence of the unqualified success of this mission of co-operation and good will. We had nominated Professor Hill with a full confidence in his personal and scientific qualifications for such an undertaking. We knew, however, that the success of his mission would depend even more on the response made to his appeal by India's men of science. To succeed, he must find them sharing his own conviction, that only a large and generous promotion of scientific and technological development in India would open for the teeming millions of its peoples any prospect of the advances in nutrition, health, prosperity and culture required to fit them for their proper place in such a world civilization as we shall soon be striving to rebuild on firmer scientific foundations. The response of India's men of science has been no more in doubt than that of its Government. We sent Professor Hill to hold out the right hand of co-operation; India has sent us six of its scientific leaders to grasp it here in Britain, and we rejoice to have them with us to-day.

In accordance with a plan which I mentioned here a year ago, the first meeting of the Society to be held outside this country, in all its long history, duly took place on 3 January of this year, on the occasion of the meeting of the Indian Science Congress in Delhi. The Viceroy, Lord Wavell, honoured the meeting with his presence, and Professor Hill, as a Vice-President, was able to give formal admission to two of our Indian Fellows, Professor Bhabha and Sir Shanti Bhatnagar, and to receive their signatures on a special sheet of parchment. India's prompt and generous response to the invitation which Professor Hill extended, has now brought Sir Shanti Bhatnagar here as a member of the return mission, so that he has been able already to sign the Charter Book itself. We shall hope that opportunities will yet occur for all Indian men of science who have been, or will in the future be, elected to our Fellowship, to come here to the Society's home and to inscribe their names in the book which our Fellows have signed since 1662. This, our ancient and now historic Charter Book is largely representative, in its earlier pages, of the springtime of modern science, not only in this country, but in all the western world; and we desire that it shall now also bear witness to a growing freedom of scientific contact and interchange with the great centre of eastern culture, from which our visitors have come. Their journey to Britain, and later to Canada and the United States of America, should mark the beginning of a new era of collaboration in science between India and the rest of the world, growing ever easier and more intimate as the means of world transport, with the added impetus which war's demands have given, attain yet higher levels of speed and safety.

Our Indian colleagues are coming to the end of a visit which, we hope, has given them a new insight into the scientific activities and organization with which this country is still meeting the demands of war, and preparing for the tasks which will later have to be shouldered by a tired though triumphant country, in a largely devastated world. We hope that they will have learned, perhaps from our failures as well as from our achievements, lessons which will have applications of real value to their own country's scientific problems. All Fellows of the Royal Society have been glad to know that our Apartments have served our Indian guests as an official home and a point of departure for their various visits and engagments, and that Professor Hill has thus been able to keep a close and friendly eye upon this gratifying fulfilment of his hopes and his plans. Two days ago this visit of our Indian colleagues, nearing its end, reached also its climax, when Their Majesties the King and Queen were graciously pleased to receive them, and thus to show their interest in the promise of a closer understanding and comradeship in science between India, the United Kingdom, and the whole of the British Empire.

I made preliminary mention last year of the action taken by the Council, in appointing a special Committee to consider the prospective needs of fundamental researches in Physics. As was then foreseen, the news of this response to an appeal from the sponsors of Physics soon evoked similar pleas from those concerned for such researches in a number of other branches of science. The Council agreed, accordingly, to appoint a series of other committees to advise upon the respective requirements, in the period following the war, of fundamental researches in Chemistry, Biology, Geology, Geophysics, Geography and Meteorology. Reports from all these Committees have been received and considered, together with one of similar aim presented by the standing Sub-Committee, on Oceanography, of the National Committee for Geodesy and Geophysics; and it is by no means certain that the tale of these special reports and appeals is entirely complete even now. It will be understood that the Society, in undertaking this investigation, has endeavoured to centre its enquiry on researches aiming at the advancement of knowledge without immediate or even implicit reference to practical needs or objectives. It has done this, not because it regards those researches, which have in view the use of scientific methods and principles for the service of man's material needs, as of inferior status or interest, or as being, by their nature, less worthy of the Society's consideration. We do not need to be reminded that researches seeking only the advancement of pure science have often revealed the shortest ways to practical developments, or that those undertaken with practical ends may chance to open windows upon new vistas of knowledge. It may be hoped that the Royal Society will never lose the comprehensive aim defined in its Charter, of 'further promoting by the authority of experiments the sciences of natural things and of useful arts'. It will be understood, however, that the Council had to keep in mind the increasing extent to which support by public money is already given, through the three Advisory Councils, to those researches which are aimed, more or less directly, at obtaining and using scientific knowledge for the promotion of industrial and material progress, of human health and of agriculture; and they had to keep also in view the more immediate appeal which is made to private benevolence and enterprise, by researches with objectives thus already in view. As I said last year, the Royal Society must, in such circumstances, regard as its special concern the proper nourishment of the roots of the tree of knowledge. We hope that the report which the Council have now adopted, based upon those of the special Committees, will enable the Society to evoke an effective response to its own conviction of the supreme importance, for the nation's future, of provision for the support of scientific researches at the frontiers of the unknown, on a scale more generous than any which has hitherto come within the range of official vision. The Society is making representations to the appropriate Departments and Authorities, and stating its conclusions as to the nature and the dimensions of the various grants of funds and facilities which will be needed, to enable this country to keep its place in the general advance of science, in the altered world which will emerge from the war.

I ought to make further brief reference to another matter on which I spoke in some detail last year, namely the accommodation provided by the nation for those scientific societies which, according to the pattern and tradition of our own, devote their activities to the promotion of natural knowledge for the general advantage, without reference to private interests, or to professional status and discipline. Such provision had been accepted as a national obligation as early as 1778. The quarters at present provided, here in Burlington House, represent the allotment made in 1869 to the societies which then achieved such recognition. Since that date no enlargement or material alteration has been made, and there seemed to be no prospect of any, either for the societies admitted 75 years ago, or for others whose claims to such privilege have later appeared. Even the accommodation thus assigned in 1869 appears to have been only a remainder, after the major part of an earlier and more generous plan for science had been diverted to other purposes by intervening claims. In 1869, we should remember, the study and promotion of the sciences had hardly been recognized as a life's occupation, except for a few determined devotees; science was at that date still largely a matter for the leisure of cultured amateurs. It is no matter for surprise, therefore, that accommodation then regarded as acceptable should now be found to make grossly inadequate provision for some of the more important of present scientific needs, and none at all for others.

After I had drawn public attention to this anomaly last year, we were asked, by several of the specialist societies concerned, to take appropriate action with the Government. We approached the Lord President of the Council, Mr Attlee, and on 13 October, with the Chancellor of the Exchequer, Sir John Anderson, and the Minister of Works and Buildings, Lord Portal, he kindly received a deputation representing all the Societies which have been housed here since 1869, and several others for which the Government has hitherto made no provision. Apart from the special and urgent needs of some of the societies, the case was fully presented for the inclusion, in any scheme for the rebuilding of London, of a centre adequate to accommodate the principal scientific societies on a well-co-ordinated plan, and so placed and designed as worthily to symbolize the central importance of British science, for the cultural as well as the material aspects of the civilization of this nation, and, indeed, of the world. The Ministers gave a careful and sympathetic hearing to the statement of our claims, and asked us to furnish quantitative data, as a basis for their further consideration of the problem.

With the outcome of the war even more certain, there can be no relaxation yet of the demand on what our scientific effort can contribute to the hastening of its end. It is none the less our duty to begin to look further ahead and to prepare for the part which science must play in the world which will follow. The needs of alliance in war have evoked, especially between the two great branches of the English-speaking nations, a closer interchange and collaboration in science, between men of different national traditions and loyalties, than has ever before been a matter of organized policy. It is not too early to begin to consider, to what degree, and in what form, such a collaborative effort should be continued into the conditions of peace, and extended to scientists of international good will throughout the world. Even in the twenty years of uneasy armistice which ended in 1939, a measure of co-operation among the world's scientists was achieved, which held promise as a potential agent of international understanding. Our traditions go back to days when our Fellows belonged to a community embracing all Europe in its enthusiasm for the new experimental philosophy, and the Royal Society will have a particular duty to be among the leaders in the resumption of international activities in science, and to use all its influence to establish these on an ever wider and firmer basis. We have a standing Committee on international relations in science, with our Foreign Secretary appropriately as its Chairman, to prepare for what action the Society can usefully undertake or promote, as the opportunity presents itself. Meanwhile, we may observe other signs that the spirit of international friendship and recognition in science is beginning to move again, even while the chaos of war is still with us. Among such signs, we may note that one of the Committees which, since their foundation in Stockholm, have awarded the Nobel Prizes with unchallenged impartiality among the scientific discoverers of all nations, has resumed its awards this year. We in this Society have welcomed the return to London of a group of distinguished French leaders in science from the United States of America, whither they had escaped from the hostile occupation of their country. We have been able to share their rejoicing at the liberation of France, and to welcome here others who had remained there, often in hiding and always in peril, as leaders in the steadfast resistance opposed by all but a negligible minority of the French men of science to the enemy's demands for their collaboration. Even today we are able to welcome another distinguished French colleague just arrived from Paris, Professor Emil Borel. The Society had the pleasure of providing a special opportunity for our Fellows to meet our French colleagues, with those from other allied countries who have long been in exile here during enemy occupation of their homelands. We are glad to think that new and lasting bonds of comradeship in science have been created for us with those who have been our country's wartime guests, and through them with all the men of science in the countries which they represent. A happy chance brings also to our meeting today four men of science from Soviet Russia. Very near to the heart of every British scientist is the desire for a growing intimacy of confidence and collaboration with our colleagues of that great partner-nation in the war for the world's freedom.

The Society had recently the opportunity of showing its interest in the revival of international scientific co-operation in another special connexion. From 16 to 19 October a small International Conference met here, under the auspices of the Health Organization of the League of Nations, to discuss the creation of an international standard of reference for Penicillin and the definition, in terms of this, of a unit of activity. Though the League has failed tragically of its central purpose, it has less conspicuous achievements to its credit, and science has an interest in ensuring the permanence of some of these. I have myself had the privilege of taking part in the activities of an International Commission under the Health Organization of the League, which succeeded, in the years between the wars, in obtaining world-wide acceptance of standards and units of activity for a whole range of modern remedies—antitoxins, hormones, vitamins and certain drugs—the strength of which could only be determined by direct biological measurements of the specific activity, in comparison with that of a fixed standard preparation in each case. Insulin was an early instance of a new remedy requiring such intervention; its general use for the treatment of diabetes could not have attained the present level of safety and effectiveness, unless a world-wide uniformity on these lines had replaced the chaos of widely different units in different countries, which was threatened in 1923. And now research has produced another new remedy, Penicillin, the success of which, in the treatment of a range of dangerous infections, has also had such a dramatic quality, that its reputation has spread rapidly beyond our scientific community and caught the interest even of a war-distracted world. Here, indeed, was a discovery which could rank as a major contribution of science to the mitigation of the suffering which war inflicts, and, at the same time and no less, as a gift of healing to mankind at peace. The needs of war had given a stimulus to the researches which proved Penicillin's remedial value, but its rapid production on an adequate scale had to face greater difficulties in our own country, where material and human resources had been more completely absorbed by earlier requisitions, than in the United States. So the present position was reached, in which, as we are proud to recognize, the existence of Penicillin and then, after a decade, the methods by which it could be separated in sufficient purity to demonstrate its brilliant possibilities as a remedy, were discoveries made here in England, while, in the further researches and technical developments needed for its large-scale production, our American colleagues have played a major part.

Thus early in its history, therefore, Penicillin and its applications had become a matter of international concern; and, though war had restricted the work in this field almost entirely to scientists of the English-speaking peoples, and had brought them into an unusual intimacy of co-operation, progress had been so rapid and action so urgent, that there was a real danger of a divergence of meaning in the terms used to express its activity and define its dosage, even among the few countries already using it. Prompt action was required to avert this by accepting a common standard of reference; and, when the proposal of a Conference for this purpose was made from this country to the League of Nations Health Section, we were grateful to our colleagues from the United States, as well as from Canada and Australia, for the generous promptitude with which they agreed to make the journey to England, so as to meet with us here in London. After all arrangements for the holding of the Conference here had been completed, the liberation of Paris opened a new possibility; Dr Tréfouel, now Director of the Pasteur Institute in Paris, was able at the last moment to accept an invitation to join us, and thus to give our deliberations, and our eventual agreement, a wider international basis. To illustrate how rapidly a divergence may arise under present conditions, I may just mention the fact that several different Penicillins have now been recognized, produced by variations in the metabolism of the growth, possibly due to mutations of the mould itself, possibly to changes in the nutritive conditions offered to it by the medium or the cultural method employed. Three such varieties of Penicillin have already been isolated in pure condition, and distinguished by certain chemical characters; but, while English workers had come to refer to these as Penicillin 1, 2, and 3, their colleagues in America spoke of Penicillin F, G, and X; and it was not until they met round the table in our Council Room a few weeks ago, but then in less than 10 minutes, that they became quite certain of the identity of 1 with F, of 2 with G, and of 3 with X. All these Penicillins have the specific remedial action in high, though not quite identical degrees, and there are probably differences, still to be explored, in their proportional efficiencies against different infective organisms. When once their identities were thus put beyond doubt, however, the small Conference had no hesitation in deciding, for the present, to use as the common basis of reference a sample of the Penicillin which is predominant in most preparations now available, and most easily obtained as a pure salt in adequate quantities. The unit could then be defined as the activity of a precise, though very small weight—0·6 μg.—of a particular sample of the perfectly dried, crystalline sodium salt of Penicillin 2, or G; and the unit thus chosen for definite fixation, and for international recognition henceforward, was, by a unanimous choice, so defined as to be as closely equivalent as possible to the unit first propounded by Sir Howard Florey's team of collaborators, and widely known as the 'Oxford' unit.

The International Standard for Penicillin is thus added to an already numerous series, of which the custody, on behalf of the League of Nations Health Organization, has been shared by our own National Institute for Medical Research with the

State Serum Institute of Denmark, at Copenhagen; and all of these standards, we may hope, will be available for transfer to whatever international authority may be established in succession to the League, as a tangible and material result of genuinely international collaboration, which the League has been able to initiate and maintain among men of science, to the permanent advantage of the world.

Though Penicillin has rightly made a special appeal to the imagination and sympathetic interest of a wide public, it is, of course, only one out of a varied range of inventions and discoveries, hastened by the stimulus of war's demands and produced, in many cases, behind the veil of its secrecy, but ready, when peace returns, to take their proper place as new gifts to the welfare and the civilized progress of mankind. From what has already been made generally known, it is clear that we may look forward to revolutionary advances in the means of communication and in the speed and safety of travel across the world, and in methods of controlling insect pests and the diseases which insects convey; and these are but a few examples of the gains which we and the world may hope to set against the tragic loss and sacrifice of the years of war. There were probably few who even suspected in 1939, that science, in countries then so dangerously unready, would find itself, before the war ended, in its present position of central importance. None of us, I think, would claim more for science even now, than to have played in this war a part of growing predominance in the provision for the fighting men of the material means of warfare, without which their heroism and sacrifice could not have prevailed. Even that duty, loyally accepted, is one from which the scientific community of the free nations must long for the release which victory will bring. But, while the operations of war have come to depend on science to a degree beyond all earlier experience, it cannot be doubted that little more than a beginning has yet been made in exploiting the possibilities of destruction, which science could progressively offer, if the world should continue thus to misuse it, and if science were still on offer for such ends. Allow me to quote a passage from a letter which the Prime Minister, whom we are proud to number among our Fellows, wrote a year ago to Professor Hill, in sending his greetings to Indian men of science.

'It is the great tragedy of our time', wrote Mr Churchill, 'that the fruits of science should by a monstrous perversion have been turned on so vast a scale to evil ends. But that is no fault of science. Science has given to this generation the means of unlimited disaster or of unlimited progress. When this war is won we shall have averted disaster. There will remain the greater task of directing knowledge lastingly towards the purposes of peace and human good.' Noble words indeed, and a profession of faith which will find an immediate echo in the hope and the desire of every true man of science. 'When this war is over we shall have averted disaster'—surely that is a confidence which every one of us will long to share. It must be clear, however, that Mr Churchill's reference was to the present threat of disaster, from which the prospect of our escape is even more fully assured to-day than when he wrote, a year ago. We may be certain that nobody sees more clearly than he, that the threat of final disaster to all man's hopes and achievements will not be forever averted, if the possibility of the 'monstrous perversion' of science is allowed to remain and to continue its evil growth. Even in the past year our enemies have thrown a new and vivid light on future possibilities, by the new weapons which science has enabled them to put on trial for our destruction. Though a people's unflinching courage and an answering effort of science and organization, together with the progress of the allied armies over the launching areas, have given us confidence that flying bombs and the like will not affect the issue of this war, the warning which they give, as to what the future might hold, is not the less clear. The writing on the wall must be plain for all to read. If, when the memories of the present war begin to fade, the world should allow science again to be exploited by a nation grasping at predominance by conquest, science will no longer be invoked only as an aid to what valour can achieve by land, sea or air, but as an agent, in itself, of blind annihilation at an ever lengthening range. When we men of science regain that freedom, for the ultimate preservation of which we have loyally accepted, through these tragic years, the bonds of secrecy and submission to authority, we cannot put aside with these our proper share in the new responsibility for the future of mankind, which this war's experiences have laid upon the men of good will in all nations. It is true, indeed, that neither the present abuse of science, nor any possibility of final disaster to civilization, which might come of a future perversion of its powers, can be charged as a fault to science itself; no more, indeed, than we could properly charge to religion, as such, the wars which once devastated much of Europe in its name. But we men of science cannot escape from our growing share in the responsibility, in 'the greater task', as Mr Churchill has written, 'of directing knowledge lastingly towards the purposes of peace and human good'. No man of science has the right to prescribe for another his interpretation in detail of that duty; but there is one aim which may unite us, perhaps for the most effective action within our common grasp, and one which is worthy of all our common influence and effort. Let me quote again from Mr Churchill's letter: 'in this task', he writes, 'the scientists of the world, united by the bond of a single purpose which overrides all bounds of race and language, can play a leading and inspiring part'. To build anew, and on a firm and broadening foundation, a world community in science, is surely an aim worthy of our utmost effort and devotion; but there can be no swerving from the present duty, and the call on science by war may yet be sterner, before we have won the freedom thus to work for the future of the world.

Awards of Medals, 1944

The COPLEY MEDAL is awarded to Sir GEOFFREY INGRAM TAYLOR in recognition of his contributions to knowledge of aerodynamics, of hydrodynamics and of the structure of metals, which have had a profound influence on the advance of physical science and of its practical applications.

Geoffrey Taylor is probably the most accomplished living exponent of the application of the methods of classical dynamics to problems of fluid motion. To great mathematical powers he adds high skill as an experimenter. His theoretical work is particularly noteworthy for its approach to reality. In place of the ideal conceptions presented by perfect incompressible fluids moving in stream-line motions and perfectly elastic solids, with which his great forerunners at Cambridge dealt, Taylor has studied turbulent motion, viscous and compressible fluids, and plastic movements of metals, obtaining results of great importance for the understanding of a wide range of phenomena.

His early work was concerned with eddy motion in the atmosphere, and opened up new fields of meteorological investigation. It threw light on the variation of wind with height and on the transference of heat and water vapour in the atmosphere, with a consequent bearing on the formation of fog. He also carried out work on the tides. Later he developed the theory of general turbulence, to which statistical methods can be applied which are somewhat reminiscent of the kinetic theory of gases.

Among Taylor's extensive researches on precise hydrodynamical problems, that on the motion of a viscous fluid between two coaxial cylinders, rotating with any speed in the same or in opposite directions, may be particularly mentioned, since it offers the only case so far of the complete solution of a problem of motional instability in the viscous liquid. In dealing with the elastic deformation of metals Taylor has shown how the slip planes can be determined in certain cases by purely geometrical methods, and has offered a formal theory of the process of work-hardening in single crystals.

Taylor has also applied his great mathematical powers to a variety of practical questions. In the last war he did work of great distinction on aerodynamical problems for the Advisory Committee on Aeronautics, and in the present war he has been extensively concerned with complicated problems concerning the propagation of explosive processes. Taylor's work may be said to be in the line of a great British tradition, which, in the past generation, was represented by investigators like W. J. M. Rankine, Osborne Reynolds and Rayleigh. Like these he has the mathematical equipment, the originality and the insight required for the fundamental solution of problems presented by practical experience in the laboratory, in the workshop, and in the wider world. Taylor has carried his quest for experience and for scientific problems on to the sea and into the air. He is a practical yachtsman and mariner, with a new design of anchors to his credit, and an aviator who, in the last war, made pioneer descents from aeroplanes by parachute. His work during this war has been of the greatest value to the nation and its allies, and his fundamental discoveries are extending the boundaries of knowledge for all mankind.

The RUMFORD MEDAL is awarded to Dr HARRY RALPH RICARDO in recognition of his important researches on the internal combustion engine.

There is a special fitness at the present time in the award of this Medal to one who, during the last twenty years, has been the leading spirit in the development of the high-speed internal combustion engine. Ricardo's researches were begun under Bertram Hopkinson in 1905 and continued, after he left Cambridge, as a consulting engineer in his grandfather's firm. Investigating the effect of turbulence on the speed of combustion, he was led to appreciate the importance of 'knocking', to determine its cause and to show that the tendency to 'knock' was dependent on the nature of the fuel. Taking charge of a special design department for his firm, he produced a four-cycle, supercharged aero-engine, long in advance of accepted practice. In 1916 he was invited to plan a special engine for the secret fighting machine which was to become known as the tank, and his unorthodox and daring design was an outstanding success. Forming a private company to maintain a laboratory for research on the internal combustion engine, Ricardo further investigated the relation of the phenomenon of 'knocking' to the maximum compression-ratio of the engine and to the character of the fuel, matching the latter by adding toluene in variable proportion to heptane, and thus paving the way for the modern octane-rating. It is not possible here to make more than general reference to the far-reaching influence of Ricardo's investigations and his steady advocacy on the designs of slide-valve engines, of sleeve-valve aero-engines, of high-speed Diesel engines, and on other important developments in engine-design. In all directions there is evidence of his special genius and flair for design, and, behind this, of his full appreciation of the thermodynamical principles which control the behaviour of engines, and of a deep knowledge of the physical and chemical factors involved, as well as of the characters of fuels and of the materials of the working parts.

A ROYAL MEDAL is awarded to Professor DAVID BRUNT in recognition of his fundamental contributions to meteorology.

David Brunt has made fundamental contributions to meteorology in its statistical, dynamical and physical aspects. The subjects which he has treated include cycles in weather; atmospheric radiation; atmospheric turbulence; the dynamical causes of rainfall; instability and convection in their bearing on the forms of clouds and on soaring flight; and the dynamics of depressions and anticyclones. He has rendered an outstanding service to his subject by his book on *Physical and Dynamical Meteorology*, in which he gives the first connected and critical account of the physics and dynamics of the atmosphere, and reduces to order a large amount of material which was previously available only in isolated papers. The book contains much original matter, and has played a leading part in the recent development of meteorology in all countries.

Brunt has been a pioneer in the analytical approach to his subject. Of recent years he has devoted several papers to the discussion of the factors which influence bodily comfort, and has gone far to provide a natural basis for the classification of climates in relation to human health and human needs.

For many years Brunt has conducted, at the Imperial College of Science and Technology, a flourishing school of meteorology, which has attracted students from all parts of the world. He has always been generous with his services to colleagues in other fields who have required expert meteorological assistance. During the war his wide knowledge and sound judgment on meteorological questions has been of the utmost value to the cause of the nation and its allies.

A **Royal Medal** is awarded to Dr **Charles Robert Harington** in recognition of his work on the structure and synthesis of thyroxine, and on the chemical basis of immune reactions.

Harington's reputation, as a leader among biochemical investigators, was established by a brilliant series of researches dealing with the chemical nature, the origin, and the form of the natural combination of the thyroid hormone, thyroxine, with its remarkable content of iodine. He improved the method of isolating this active principle from the thyroid gland, determined its structural constitution and then produced it by artificial synthesis. Later he demonstrated that diiodotyrosine was present in the gland and accounted for the balance of its iodine content. By enzymatic cleavage he proceeded to show that thyroxine and diiodotyrosine are natural amino-acid constituents of the complex thyreoglobulin. These discoveries gave an entirely new precision to knowledge of the thyroid hormone, of the manner of its natural occurrence and function, and of the diseases which result from excess or defect of its supply from the gland. In more recent years Harington's work has contributed very important advances to knowledge of the chemical basis of immunological specificity. By a new method of coupling haptene groups artificially to proteins, he has studied the role of carbohydrates and of tyrosine in developing antigenic properties. By such methods he has created artificial antigens, the specificity of which is determined by the attachment of physiological active haptenes, such as thyroxine and acetylsalicylic acid. He has thus produced and determined the limits of specificity of antisera reacting with free thyroxine or acetylsalicylic acid, and has observed the antagonism of such sera to the physiological effects of such principles in the animal body. In other directions also Harington has made brilliant contributions to biochemical knowledge, as by his work on the conditions determining the crystallization of insulin and on the synthesis of glutathione.

The DAVY MEDAL is awarded to Sir ROBERT ROBERTSON in recognition of his researches on explosives, analytical methods, the internal structure of the diamond and infra-red absorption spectra.

After studying at St Andrews, Robertson made his first acquaintance with the field of explosives as a chemist at the Waltham Abbey Royal Gunpowder Factory, where he was occupied on the nitroglycerine plant. He acquired a knowledge of all aspects of the manufacture of cordite and contributed improvements, such as the acetone recovery process which has been widely adopted. His researches in this period covered calorimetric measurements and in particular the study of the kinetics of the decomposition of gun cotton, which led to the publication of work of great practical importance on the stabilization of that material (1916).

A succession of spontaneous explosions in cordite magazines led to a visit by Robertson to India and to the issue of an exhaustive and valuable report in February, 1907. Shortly afterwards he entered the Research Department at Woolwich as Superintending Chemist, later director of Explosives Research, and he occupied these posts with distinction until 1920. Much valuable work was done before and during the war of 1914–18. A process for the manufacture of T.N.T. introduced novel features and prepared the way for the large T.N.T. factories established under Lord Moulton. Other notable achievements by Robertson were the introduction of cordite R.D.B., which relieved the acetone position, and of amatol, of which it was said by the Director of Artillery that 'amatol won the war'. Throughout his time at the Woolwich Research Department, Robertson showed himself a resourceful investigator, an able leader and an indefatigable worker.

There followed for Robertson a further period of effective organization and active enterprise in research, as Government Chemist. He was one of the first to recognize the importance of infra-red spectrography for the determination of molecular structure, and in his pioneering work on the infra-red spectra of ammonia and of arsine he pushed the accuracy of the instruments then available to their ultimate limits. These researches drew the attention of chemists to the possibilities of the analysis of molecular vibrational and rotational bands, and materially assisted in opening up the wide field which has been explored in recent years. Robertson's studies of the absorption spectra of diamonds have produced results of very great interest; they show that diamonds exist in two types differentiated by the condition of strain originating in their high-temperature formation, and by the mosaic character of the less usual type.

During the present war Robertson has occupied very responsible positions in relation to the earlier field of his researches and has played a keen and active part in the contribution of chemistry to the national emergency.

The DARWIN MEDAL is awarded to Dr JOHN STANLEY GARDINER in recognition of his life's work on coral reefs.

Gardiner is universally recognized as an authority on coral reefs and on the organisms associated with such habitats. His contributions to these fields of biological and geographical research began not long after his graduation, when he was a member of the coral reef boring expedition to the Atoll of Funafuti, organized by the Royal Society in 1896. Since then he has himself organized and led two most important expeditions, the first to the Maldive and Laccadive Archipelagoes in 1899 and the second to the Indian Ocean in 1905; the results of these expeditions are embodied in nine large quarto volumes and represent a most valuable contribution to a field of knowledge closely associated with the work of Charles Darwin. Within recent years Gardiner has organized and, to a large extent, directed the Cambridge Expedition to the Suez Canal, 1924; the Great Barrier Reef Expedition, 1928–31; the John Murray Expedition to the Indian Ocean, 1933–34; and the Expedition to Lake Titicaca in 1937. He is an authority on the taxonomy and systematics of Alcyonarian and Zoantharian Corals, and has taken a keen interest in their ecology and geographical distribution. He realized the great importance, in the study of corals, of the examination of the polyps themselves, as well as of their coralla, and he paid special attention to variations which may result from slight differences of habitat and are correlated with physical and other conditions, showing that in several instances so-called 'species' are merely variations.

He also realized the immense value of an accurate knowledge of the coral fauna of any given locality in relation to its environment, in enabling one to deduce the conditions under which tertiary and earlier coralline deposits have been formed.

There is hardly a branch of research on corals and coral reefs in which Gardiner's work is not of great importance. It was his observations on the Funafuti Atoll and the Atolls of the Maldive and Laccadive Archipelagoes that caused him to realize that no one theory, such as the 'subsidence' theory of Darwin, or the 'solution' theory of Murray, can account for the formation of all such reefs and atolls though, when once formed, every reef has been moulded and modified by world-wide phenomena, such as a change in the relative levels of sea and land.

Stanley Gardiner has given us an admirable summary of this, his life's work, and of the conclusions that he has drawn from it, in his book *Coral Reefs and Atolls*, a most valuable supplement to Darwin's own volume *On the structure and distribution of coral reefs*. There is a special fitness in the award of the Darwin Medal for work of such 'acknowledged distinction in the field in which Charles Darwin himself laboured'.

The HUGHES MEDAL is awarded to Professor GEORGE INGLE FINCH in recognition of his fundamental contributions to the study of the structure and properties of surfaces; and for his important work on the electrical ignition of gases.

Finch has carried out two important bodies of research in different fields, both involving electrical considerations in a fundamental manner. The first was a detailed study of the electrical ignition of gases, the second the application of electron diffraction to a wide range of chemical and physical surface problems.

In his work on electrical ignition, he not only elucidated the chemistry of the ignition of simple gaseous systems, but was the first to develop the theory of the sparking ignition coil. His inductance component control interrupter has been used by the Radio Research Board for the production of single electromagnetic pulses. In the course of his ignition work Finch developed, as a pioneer in this country, the high speed cathode-ray oscillograph.

In the field of electron diffraction Finch has developed the electron diffraction camera into an equipment giving results of high accuracy with speed and ease of manipulation. The Finch camera has found wide application outside his laboratory and examples made under his direction have been installed, among other places, at the National Physical Laboratory, University College, London, the University of Brussels (two), and in the laboratories of Messrs Ferranti and of other industrial research centres. The pictures which he has obtained with it are outstanding in beauty of detail. He has contributed notably to the interpretation of the electron diffraction pattern and has applied his methods to many problems of theoretical and practical importance.

Of special interest are his studies of the relation between crystal size and lattice dimensions and, in the more practical field, his investigations into the effect of the substrate on adhesion of electro deposits, into the nature of polish and into the mechanism of boundary lubrication and the wear of sliding surfaces. In all these he has materially advanced our knowledge, and his work on sliding surfaces, in particular, has found important applications in engineering practice.

His work during the war has covered a variety of fields, some involving the application of electron diffraction. That which he has carried out as scientific adviser to the Ministry of Home Security, while less closely related to his normal lines of research, has been of the greatest value.

Relaxation methods applied to engineering problems
XI. Problems governed by the 'quasi-plane-potential equation'

By D. N. de G. Allen, R. V. Southwell, F.R.S. and Gillian Vaisey

(*Received* 24 *March* 1944)

The 'quasi-plane-potential equation'

$$\frac{\partial}{\partial x}\left(\chi\frac{\partial\psi}{\partial x}\right)+\frac{\partial}{\partial y}\left(\chi\frac{\partial\psi}{\partial y}\right)+Z=0$$

is discussed in relation to problems where Z and χ are *known* functions of x and y. It governs (*inter alia*) the small transverse displacement of a membrane in which the tension $T \propto \chi$, and its finite-difference approximation governs the small transverse displacements of nodal points of a net in which, similarly, the string tension T varies from node to node. (Equilibrium in the directions of x and y can be maintained, both in the membrane and in the net, by forces acting in those directions and accordingly having no effect on the transverse equilibrium.)

The relaxational treatment, based on this mechanical (net) analogue, reduces when χ is constant to the treatment developed, in earlier papers of this series, for problems governed by the 'plane-harmonic' (Poisson) equation

$$\frac{\partial^2\psi}{\partial x^2}+\frac{\partial^2\psi}{\partial y^2}+Z=0.$$

Thus an opportunity is afforded for a review of various improvements which have been incorporated in the technique first propounded in Part III. In particular, this paper for the first time uses *systematically* the device of 'graded nets', i.e. nets of which the mesh-side a is smaller in some parts than in others.

Introduction

1. Relaxation methods as developed in this series are direct extensions of a method first suggested for the calculation of stresses in frameworks—'the method of systematic relaxation of constraints' (Southwell 1935). Part I brought differential equations within their scope. Part III first dealt with equations involving two independent variables. Part VI developed a technique for 'characteristic value' (*eigenwerte*) problems.

Part VII and papers subsequent thereto* have all been concerned with equations in two variables. Like Part III (which dealt with the equations of Laplace and of Poisson) they begin by substituting for the exact differential equation its approximation in finite differences, thereby obtaining (in effect) a system of equations relating values of the wanted function at nodal points of a regular lattice or 'net', then they deal with those equations by a relaxational technique. Successive papers have treated governing equations of increasing order and complexity, or boundary conditions of special types; but from time to time they have recorded improvements of technique which have general application, though suggested by special problems.

---

* Some of these papers have been adjudged to be 'secret' during the war period, and accordingly are not included in the list of references.

It was inevitable that relaxation methods should thus develop, but a consequence is that no paper in the series gives an exact account of those methods as now employed; Part III in some respects is superseded, and later papers are concerned with particular applications. In this paper, therefore, we review the present state of relaxational technique as a preliminary to particular study of the equation

$$\frac{\partial}{\partial x}\left(\chi\frac{\partial\psi}{\partial x}\right)+\frac{\partial}{\partial y}\left(\chi\frac{\partial\psi}{\partial y}\right)+Z=0 \tag{1}$$

in relation to problems where χ and Z are known functions of x and y.* When χ is constant, (1) reduces to the 'plane-harmonic equation' of Parts III, V, VII and VIII, and on this account we have ventured to describe its form as 'quasi-plane-harmonic'.

BASIC THEORY

2. When Z is zero, (1) reduces to

$$\frac{\partial}{\partial x}\left(\chi\frac{\partial\psi}{\partial x}\right)+\frac{\partial}{\partial y}\left(\chi\frac{\partial\psi}{\partial y}\right)=0, \tag{2}$$

and the existence of a solution ψ implies that

$$\chi\frac{\partial\psi}{\partial y}=\frac{\partial\phi}{\partial x},\quad -\chi\frac{\partial\psi}{\partial x}=\frac{\partial\phi}{\partial y}; \tag{3}$$

therefore implies the existence of a solution ϕ to the equation

$$\frac{\partial}{\partial x}\left(\chi^{-1}\frac{\partial\phi}{\partial x}\right)+\frac{\partial}{\partial y}\left(\chi^{-1}\frac{\partial\phi}{\partial y}\right)=0 \tag{4}$$

which results when ψ is eliminated from (3). When χ is constant it can be cancelled out from (2) and (4), and then both ψ and ϕ (its 'conjugate') are plane-harmonic functions: this is a familiar result. In general the associated functions ψ and ϕ satisfy similar but not identical equations, described as 'adjoint'.

When s and ν, measured along and perpendicular to a closed contour or boundary, have the senses shown in figure 1, then

$$\left.\begin{aligned}\frac{\partial}{\partial s}&\equiv\cos(x,\nu)\frac{\partial}{\partial y}-\sin(x,\nu)\frac{\partial}{\partial x},\\ \frac{\partial}{\partial \nu}&\equiv\cos(x,\nu)\frac{\partial}{\partial x}+\sin(x,\nu)\frac{\partial}{\partial y},\end{aligned}\right\} \tag{5}$$

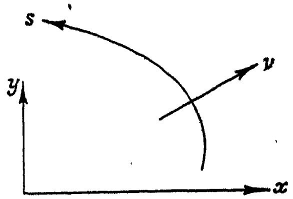


FIGURE 1

and hence, when (3) are satisfied,

$$\chi\frac{\partial\psi}{\partial s}=\frac{\partial\phi}{\partial \nu},\quad -\chi\frac{\partial\psi}{\partial \nu}=\frac{\partial\phi}{\partial s}. \tag{6}$$

This means that $\partial\psi/\partial\nu=0$ at every point of a ϕ-contour ($\phi=$ const.), $\partial\phi/\partial\nu=0$ at every point of a ψ-contour ($\psi=$ const.). These too are generalizations of familiar properties of conjugate plane-harmonic functions.

---

* Part IX discussed an equation of this form (Z being zero); but there χ was related with ψ by a second equation (the 'equation of state'). Cf. also Christopherson (1941): 'Relaxation methods applied to problems of film lubrication.'

3. When $Z \neq 0$ in (1), an adjoint equation can still be formulated. Let Ω be a function such that

$$\frac{\partial^2\Omega}{\partial x^2}+\frac{\partial^2\Omega}{\partial y^2} = Z(x,y). \tag{7}$$

Then (1), written in the form

$$\frac{\partial}{\partial x}\left(\chi\frac{\partial\psi}{\partial x}+\frac{\partial\Omega}{\partial x}\right)+\frac{\partial}{\partial y}\left(\chi\frac{\partial\psi}{\partial y}+\frac{\partial\Omega}{\partial y}\right) = 0,$$

shows that

$$\chi\frac{\partial\psi}{\partial y}+\frac{\partial\Omega}{\partial y} = \frac{\partial\phi}{\partial x}, \quad -\left(\chi\frac{\partial\psi}{\partial x}+\frac{\partial\Omega}{\partial x}\right) = \frac{\partial\phi}{\partial y}, \tag{8}$$

and from these relations the equation

$$\frac{\partial}{\partial x}\left(\chi^{-1}\frac{\partial\phi}{\partial x}\right)+\frac{\partial}{\partial y}\left(\chi^{-1}\frac{\partial\phi}{\partial y}\right)+\chi^{-2}\left(\frac{\partial\chi}{\partial x}\frac{\partial\Omega}{\partial y}-\frac{\partial\chi}{\partial y}\frac{\partial\Omega}{\partial x}\right) = 0 \tag{9}$$

is obtained by elimination of ψ. Since χ is specified and Ω is deducible from (7), the last term on the left of (9) may be regarded as a *known* function of x and y, corresponding with Z in (1).

SOME PHYSICAL EXAMPLES

4. The quasi-plane-harmonic equation occurs in many branches of mathematical physics. In electricity, for example, it governs the flow of current, and the electric potential, in a plane conducting sheet. Let i_x, i_y denote line-intensities of current in the directions Ox, Oy, and suppose that both surfaces of the sheet are insulated so that no current, on the whole, can enter or leave an element. This supposition may be stated in the equation

$$\frac{\partial}{\partial x}i_x+\frac{\partial}{\partial y}i_y = 0, \tag{i}$$

which shows that i_x, i_y can be expressed in terms of a single 'current function' ψ, thus:

$$i_x = \frac{\partial\psi}{\partial y}, \quad i_y = -\frac{\partial\psi}{\partial x}. \tag{ii}$$

Let ρ denote the 'resistivity' of the conducting sheet, V the electric potential at the point x, y. Then, by Ohm's law,

$$-\frac{\partial V}{\partial x} = \rho i_x = \rho\frac{\partial\psi}{\partial y}, \quad -\frac{\partial V}{\partial y} = \rho i_y = -\rho\frac{\partial\psi}{\partial x}, \tag{iii}$$

and elimination of V and ψ, in turn, from these equations leads to

$$\left.\begin{aligned}\frac{\partial}{\partial x}\left(\rho\frac{\partial\psi}{\partial x}\right)+\frac{\partial}{\partial y}\left(\rho\frac{\partial\psi}{\partial y}\right) = 0\\ \text{and to}\quad \frac{\partial}{\partial x}\left(\rho^{-1}\frac{\partial V}{\partial x}\right)+\frac{\partial}{\partial y}\left(\rho^{-1}\frac{\partial V}{\partial y}\right) = 0.\end{aligned}\right\} \tag{10}$$

Equations (10) are analogous with (2) and (4), for which, accordingly, they suggest a physical interpretation: The function χ, in (2), is a measure of some kind of 'resistance' (ρ) or 'admittance' (ρ^{-1}) which varies from point to point of the field. When it measures a resistance, the wanted function measures something having the nature of a flux; when an admittance, something having the nature of potential. Similarly in the theory of magnetism, if Ω denotes the 'magnetic potential', μ the 'permeability' of the material at any point (x, y, z) in a magnetic field, when everything is independent of z $\left(\text{so that } \frac{\partial \Omega}{\partial z} = \frac{\partial \mu}{\partial z} = 0\right)$, then Ω satisfies the equation

$$\frac{\partial}{\partial x}\left(\mu \frac{\partial \Omega}{\partial x}\right) + \frac{\partial}{\partial y}\left(\mu \frac{\partial \Omega}{\partial y}\right) = 0, \tag{11}$$

μ being (in general) a specified function of x and y. With (11) is associated the adjoint equation

$$\frac{\partial}{\partial x}\left(\mu^{-1} \frac{\partial \psi}{\partial x}\right) + \frac{\partial}{\partial y}\left(\mu^{-1} \frac{\partial \psi}{\partial y}\right) = 0, \tag{12}$$

which governs the 'magnetic flux function' ψ; and (12) and (11) reduce respectively to (2) and (4) when μ^{-1} (a quantity having the nature of resistance) is replaced by χ, and Ω (a potential function) by ϕ.

5. In systems characterized by symmetry with respect to an axis Oz, the quasi-plane-potential equation appears as a special form of Laplace's equation. Thus in material of uniform thermal conductivity the temperature v, if steady, is governed by

$$\nabla^2 v = 0, \tag{13}$$

which in cylindrical co-ordinates r, θ, z assumes the form

$$\left[\frac{\partial}{\partial r^2} + \frac{1}{r}\frac{\partial}{\partial r} + \frac{\partial^2}{\partial z^2}\right] v = 0, \tag{14}$$

i.e.

$$\frac{\partial}{\partial r}\left(r \frac{\partial v}{\partial r}\right) + \frac{\partial}{\partial z}\left(r \frac{\partial v}{\partial z}\right) = 0, \tag{15}$$

when (on account of axial symmetry) v is invariant with respect to θ. Equation (15) is identical in form with (2), v and r replacing ψ and χ.

6. A last example may be taken from the theory of torsional stresses in a bar of circular section but *non-uniform diameter* (Southwell 1942, §§ 8 and 11). On certain assumptions (akin to those made in Saint Venant's theory of torsion for prismatic bars of non-circular section), these consist solely of shear stress on all axial (z, r) planes, having components

$$\widehat{\theta z} = -\frac{1}{r^2}\frac{\partial \phi}{\partial r}, \quad \widehat{r\theta} = \frac{1}{r^2}\frac{\partial \phi}{\partial z}, \tag{16}$$

where ϕ is a function of r and z, satisfying the condition

$$\phi = \text{const.} \tag{17}$$

along a generator of the (stress-free) boundary surface, and the equation

$$\frac{\partial}{\partial r}\left(\frac{1}{r^3}\frac{\partial\phi}{\partial r}\right)+\frac{\partial}{\partial z}\left(\frac{1}{r^3}\frac{\partial\phi}{\partial z}\right)=0 \tag{18}$$

at every point in the material. Equation (18) is (2) with

$$\chi=r^{-3},\quad \psi=\phi. \tag{19}$$

MEMBRANE ANALOGUE OF THE QUASI-PLANE-POTENTIAL EQUATION

7. A 'membrane analogue' of (1) can be stated. Suppose that the membrane tension T has an intensity, $\mathsf{T}\chi$, which varies from point to point, and let ψ denote the (small) transverse displacement: then it is easy to obtain (1) as the condition of transverse equilibrium for an element at (x, y) under external transverse loading of surface intensity $\mathsf{T}Z$. There is no difficulty in the notion of a variable tension, since we can suppose this to be maintained by forces acting always in directions parallel to the plane of the undeflected membrane, and as such having no influence on the *transverse* equilibrium. The surface intensity of these forces must have components

$$X=-\mathsf{T}\frac{\partial\chi}{\partial x},\quad Y=-\mathsf{T}\frac{\partial\chi}{\partial y}, \tag{20}$$

so they will be derivable from a potential $\mathsf{T}\chi$.

THE FINITE-DIFFERENCE APPROXIMATION, AND THE 'RELAXATION PATTERN'

8. A relaxational treatment of (1)—which includes (2) as a special case—presents no difficulty. *It is assumed that the equation is in 'non-dimensional' form* (so that x, y, χ and ψ are purely numerical quantities), and as a first step its finite-difference approximation is formulated.

Referring to figure 2, we have the usual approximations:

$$\left.\begin{aligned} a\left(\frac{\partial\psi}{\partial x}\right)_{\mathrm{I}} &\approx \psi_1-\psi_0,\\ a\left(\frac{\partial\psi}{\partial x}\right)_{\mathrm{III}} &\approx \psi_0-\psi_3,\end{aligned}\right\} \tag{i}$$

therefore (similarly)

$$\begin{aligned} a^2\left[\frac{\partial}{\partial x}\left(\chi\frac{\partial\psi}{\partial x}\right)\right]_0 &\approx a\left(\chi\frac{\partial\psi}{\partial x}\right)_{\mathrm{I}}-a\left(\chi\frac{\partial\psi}{\partial x}\right)_{\mathrm{III}}\\ &\approx \chi_{\mathrm{I}}(\psi_1-\psi_0)-\chi_{\mathrm{III}}(\psi_0-\psi_3). \end{aligned} \tag{ii}$$

By a like argument we have

$$a^2\left[\frac{\partial}{\partial y}\left(\chi\frac{\partial\psi}{\partial y}\right)\right]_0 \approx \chi_{\mathrm{II}}(\psi_2-\psi_0)-\chi_{\mathrm{IV}}(\psi_0-\psi_4), \tag{iii}$$

and hence, as the required approximation to (1),

$$\left.\begin{aligned} \mathbf{F}_0 &= \mathsf{F}_0 + F_0 = 0, \\ \text{where} \quad \mathsf{F}_0 &= a^2 Z_0, \quad F_0 = \Sigma_4(\chi_{\mathrm{I}} \cdot \psi_1) - \psi_0 \cdot \Sigma_4(\chi_{\mathrm{I}}). \end{aligned}\right\} \tag{21}$$

In these equations I, II, III, IV (for which χ-values must be known) are points half-way between 0 and 1, 2, 3, 4 respectively. It is an easy matter to deduce from (21) the 'relaxation pattern' corresponding with an isolated displacement $\Delta\psi_0 = 1$. This is shown in figure 3*a*: it is very simple, and it reduces (as it should) to the standard plane-harmonic pattern when $\chi = 1$.

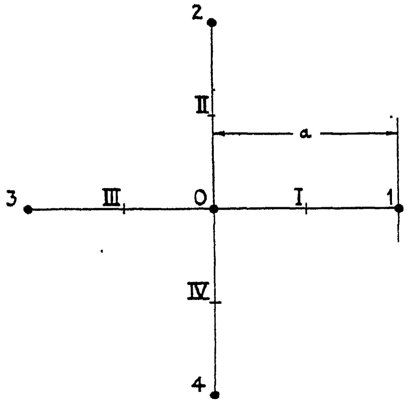


FIGURE 2

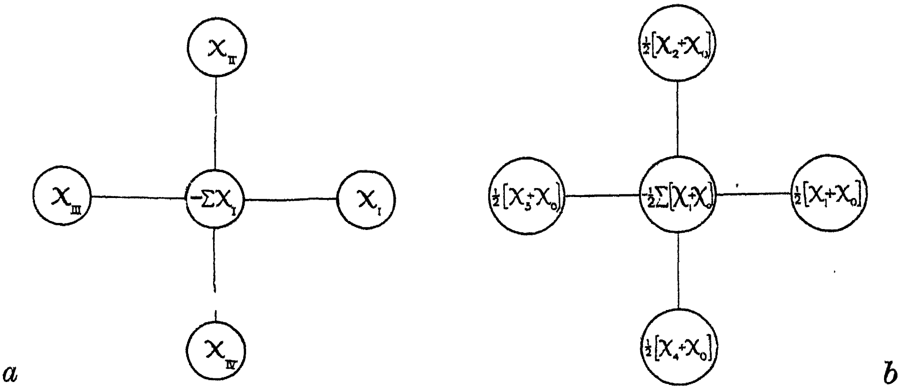

χ_{II}
χ_{III}
$-\Sigma\chi_{\mathrm{I}}$
χ_{I}
χ_{IV}
a
$\frac{1}{2}[\chi_2+\chi_0]$
$\frac{1}{2}[\chi_3+\chi_0]$
$-\frac{1}{2}\Sigma[\chi_1+\chi_0]$
$\frac{1}{2}[\chi_1+\chi_0]$
$\frac{1}{2}[\chi_4+\chi_0]$
b

FIGURE 3

9. Alternatively, since the quantity

$$\begin{aligned} 2\left\{\frac{\partial}{\partial x}\left(\chi\frac{\partial\psi}{\partial x}\right) + \frac{\partial}{\partial y}\left(\chi\frac{\partial\psi}{\partial y}\right)\right\} &= 2\left\{\chi\nabla^2\psi + \frac{\partial\chi}{\partial x}\frac{\partial\psi}{\partial x} + \frac{\partial\chi}{\partial y}\frac{\partial\psi}{\partial y}\right\} \\ &= 2\chi\nabla^2\psi + \nabla^2(\chi\psi) - \chi\nabla^2\psi - \psi\nabla^2\chi \\ &= \nabla^2(\chi\psi) + \chi\nabla^2\psi - \psi\nabla^2\chi, \end{aligned} \tag{i}$$

therefore using the standard (square-net) approximation

$$a^2(\nabla^2 w)_0 \approx \Sigma_{a,4}(w) - 4w_0 \tag{ii}$$

in relation to the terms of the right of (i), we have

$$\begin{aligned} 2F_0 &= \Sigma_{a,4}(\chi\psi) - 4\chi_0\psi_0 + \chi_0\{\Sigma_{a,4}(\psi) - 4\psi_0\} - \psi_0\{\Sigma_{a,4}(\chi) - 4\chi_0\} \\ &= \Sigma_4[\psi_1(\chi_1 + \chi_0)] - \psi_0 \,.\, \Sigma_4(\chi_1 + \chi_0) \end{aligned} \tag{22}$$

in replacement of the last of (21). The corresponding pattern is as shown in figure 3*b*, reducing to identity with figure 3*a* when

$$2(\chi_{\mathrm{I}}, \chi_{\mathrm{II}}, \chi_{\mathrm{III}}, \chi_{\mathrm{IV}}) \quad \text{is identified with} \quad \chi_0 + (\chi_1, \chi_2, \chi_3, \chi_4),$$

and to identity with the standard plane-harmonic pattern when $\chi = 1$.

Within the approximation of our finite-difference expressions for $\partial\psi/\partial x$, etc., figures 3*a* and 3*b* are equivalent: that is to say, their differences are of a fractional order which is here neglected. An advantage of the treatment leading to (22) is that it can be applied as well to triangular nets ($N = 6$) if (ii) is replaced by its generalization

$$N\frac{a^2}{4}(\nabla^2 w)_0 \approx \Sigma_{a,N}(w) - Nw_0, \tag{iii}$$

in which (cf. Part III, § 8) terms of order a^4 are suppressed. Substituting from (i) and (iii) in (1), we obtain the first of (21) with expressions as under for F_0, F_0:

$$\mathsf{F}_0 = N\frac{a^2}{4}Z_0, \quad 2F_0 = \Sigma_N[\psi_1(\chi_1 + \chi_0)] - \psi_0\Sigma_N(\chi_1 + \chi_0). \tag{23}$$

The second of these reduces to (22) when $N = 4$; and with an error of order (a^4) which is here neglected it can be replaced by

$$F_0 = \Sigma_N(\chi_{\mathrm{I}}\psi_1) - \psi_0\Sigma_N(\chi_{\mathrm{I}}), \tag{24}$$

which, when $N = 4$, reduces similarly to the last of (21). Diagrams similar to figures 3 *a*, *b*, are easily constructed to give, for $N = 6$, the relaxation patterns corresponding with (24) and (23) respectively.

NET ANALOGUE OF THE FINITE-DIFFERENCE APPROXIMATION

10. The results of §§ 8–9 can be interpreted mechanically, as relating to a tensioned net in which the string tensions (like the membrane tension in § 7) have different values in different parts; equilibrium in the plane of the undeflected net being maintained (as in § 7) by an application of suitable forces which are without effect on the transverse equilibrium.

What follows is a treatment similar to that of Part VIII, §§ 6–10, but modified in that now the membrane tension is given by

$$T = \mathsf{T}\chi, \tag{25}$$

T being constant but χ a specified function of x and y (§ 7). When $N = 6$ the loading $\mathsf{T}Z$ is concentrated at nodal points according to the rules of statics; but when $N = 3$ or 4 it is best to concentrate at O (figure 4) the whole of the pressure acting on the surrounding polygon $abc \ldots a$. Then, for $N = 6$, the initial force at O is given by

$$\mathsf{F}_0 = \tfrac{1}{4}N\mathsf{T}\tan\frac{\pi}{N}\left[a^2Z_0+\frac{a^4}{16}(\nabla^2Z)_0\right]; \tag{i}$$

when $N = 3$ or 4 it is given by

$$\mathsf{F}_0 = \tfrac{1}{4}N\mathsf{T}\tan\frac{\pi}{N}a^2Z_0, \quad \text{simply.} \tag{ii}$$

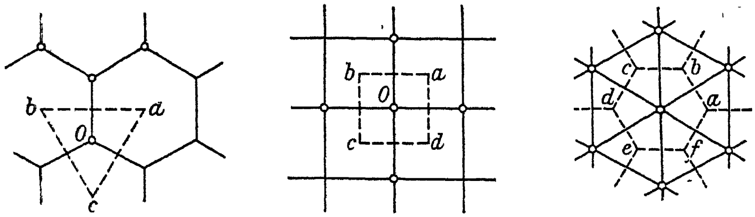


FIGURE 4

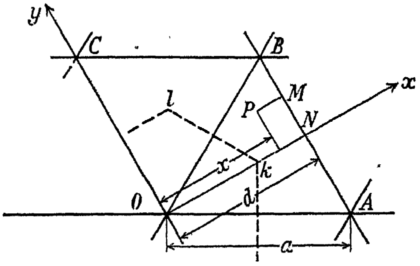


FIGURE 5

As in Part VIII, § 8, we concentrate in a string OB of the substituted net (figure 5) the whole force which is the membrane tension T acting on the length $kl = a\tan\pi/N$. Then, with neglect of quantities of order a^4,

$$\mathsf{T} = \mathsf{T}a\tan\frac{\pi}{N}\chi_{\mathrm{I}}, \tag{iii}$$

I denoting (as before) the middle point of the string: so we have the expression

$$\mathsf{T}\Delta\psi/a = \mathsf{T}\tan\frac{\pi}{N}\chi_{\mathrm{I}}\Delta\psi \tag{iv}$$

for the force exerted by a string in consequence of a relative transverse displacement $\Delta\psi$ of its two ends; and hence the expression

$$F_0 = \mathsf{T}\tan\frac{\pi}{N}\Sigma_N\,[\chi_{\mathrm{I}}(\psi_1-\psi_0)] \tag{v}$$

for the total transverse force imposed at O by reason of displacements $\psi_0, \psi_1, \ldots, \psi_N$.

11. Accordingly the condition of transverse equilibrium is

$$\mathbf{F}_0 = \mathsf{F}_0+F_0 = 0, \tag{21) \textit{bis}$$

where F_0 and F_0 are given by (i) or (ii) and (v). But this condition is unchanged if F_0, F_0 are multiplied in the same proportion, and moreover, since terms of order a^3 have been neglected in obtaining (iv), there is no longer any justification in retaining the second term on the right of (i): therefore, cancelling (or equating to unity) the factor $\mathsf{T}\tan\pi/N$, we now associate with (21) the expressions

$$\left.\begin{aligned} \mathsf{F}_0 &= \tfrac{1}{4}Na^2Z_0, \\ F_0 &= \Sigma_N[\chi_{\mathrm{I}}(\psi_1-\psi_0)] = \Sigma_N(\chi_{\mathrm{I}}\psi_1)-\psi_0\Sigma_N(\chi_{\mathrm{I}}), \end{aligned}\right\} \tag{26}$$

which, for $N = 4$, are identical with the second and third of (21), § 8.

The 'force per string' must be modified accordingly: i.e. in place of (iv) we now have

$$\mathbf{T}\Delta\psi/a = \chi_{\mathrm{I}}\Delta\psi,$$

and in place of (iii)

$$\mathbf{T} = a\chi_{\mathrm{I}}. \tag{27}$$

Because, with neglect of terms of order a^2,

$$2\chi_{\mathrm{I}} = \chi_1+\chi_0, \ \ldots, \text{ etc.},$$

with neglect of terms of order a^3 the last of (26) is equivalent to (22), and (27) is equivalent to

$$2\mathbf{T} = a(\chi_0+\chi_1). \tag{28}$$

Cancellation of the factor $\mathsf{T}\tan\pi/N$ is equivalent to the assumption that

$$\mathsf{T} = \cot\pi/N. \tag{29}$$

This compares with the assumption (14) of Part VIII, § 9, where, in consequence, the string tension $\mathbf{T}$ had a uniform value a, with the result that unit force was transmitted by any string by reason of a unit difference in the displacements of its two ends. (The same result is obtained by making χ_0 and χ_1 unity in (28). In the original treatment of Part III, § 18, the 'force per string' was made less in the ratio $1/N$, and a similar reduction was made in the expression for F_0. The new convention has been found to have practical advantages.)

'IRREGULAR STARS', AND THE USE OF 'FICTITIOUS NODES'

12. To summarize: equation (21) with F_0, F_0 given by (23) or (26)—which are equivalent—may be used in accordance with the net analogy to solve the quasi-plane-potential equation. The corresponding 'patterns' are shown in figures 3*a, b*, § 8, and may be used as in previous papers of the series to deduce 'block' and 'group displacements'. The technique of 'liquidation' need not be described again.

Figures 3 express the assumption that the tension in any string is

$$\mathrm{T} = a\chi_{\mathrm{I}}, \tag{27 bis}$$

$$= \tfrac{1}{2}a(\chi_0+\chi_1) \tag{28 bis}$$

when the third and higher powers of a are neglected. When $\chi = 1$ everywhere, both of these expressions reduce to

$$\mathrm{T} = a, \tag{30}$$

as assumed in Part VIII, § 9. There, as a result, unit force was transmitted by any one string in consequence of a unit difference in the displacements of its two ends. According to (27) and (28) the force so transmitted is $\chi_{\mathrm{I}} = \frac{1}{2}(\chi_0+\chi_1)$.

13. No difficulty is presented by 'irregular stars' (i.e. by nodal points from which one or more strings radiate which are not of the standard length a), provided that we accept the consequences of the mechanical (net) analogue and assume that every string remains straight in the deflected net. The force transmitted by any string is proportional to its tension and to its slope, and the latter, for the same difference between its end displacements, is increased in the ratio $1/\xi$ when its length is ξa instead of a. Therefore the 'force per unit displacement' is increased to

$$\frac{1}{\xi}\chi_{\mathrm{I}} = \frac{1}{2\xi}(\chi_0+\chi_1), \tag{31}$$

or to $1/\xi$ when $\chi = 1$ as in Part III, 0, 1, I denoting as before the ends and middle point of the string in question.

In cases where the wanted function has specified *boundary values* it has become customary to employ the notion of 'fictitious points' lying outside the boundary and connected with internal nodes by strings of the standard mesh-length a. Then, if unit displacement is imposed upon the internal node *while boundary points are held fixed*, to keep the string straight a displacement $-(1-\xi)\,a/\xi$ must be imposed at the same time on the fictitious node, ξa denoting as before the length of string inside the boundary. This means that a *differential* displacement a/ξ is imposed upon the two ends of a string of standard length a, and the expressions (31) result.

'Fictitious nodes' have been found useful in dealing with problems involving *refraction*, where a string on crossing an interface between two different media has to have its gradient multiplied in a specified ratio. This problem was discussed in Part III, § 34, with a result which *in our present notation* may be stated as follows: The discontinuous change of medium implies (according to the 'net analogue') a discontinuous change of string tension from T_1 (say) to T_2. Let $\alpha_1 a$ be the length of string which has tension T_1, $\alpha_2 a$ the length which has tension T_2 (so that $\alpha_1+\alpha_2 = 1$). Then the effective length of string in medium 1 is $\xi_1 a$, and the effective length of string in medium 2 is $\xi_2 a$, where

$$\xi_1 = \alpha_1+\frac{T_1}{T_2}\alpha_2, \quad \xi_2 = \alpha_2+\frac{T_2}{T_1}\alpha_1, \tag{32}$$

These formulae are to be combined with (31).

USE OF 'GRADED NETS': (1) WHEN BOUNDARY VALUES ARE SPECIFIED

14. As liquidation approaches finality, more accurate but less simple formulae can be used in place of (31) and (32), which have been derived on the understanding (cf. § 13) that strings remain straight. But greater labour is entailed by this procedure, and in general it would seem better to retain the simpler formulae and to rely for close approximation on the fineness of the ultimate net.

The use of 'successive nets' has been exemplified in several papers in this series, but there is an alternative which has not, so far, been used systematically: The accuracy of our finite-difference approximation depends upon the magnitudes of the neglected differentials, i.e. upon the rapidity with which the wanted function varies; consequently *a finer net is needed in some parts of a 'field' than in others, to define the function with precision*. In hydrodynamic problems, for example, a fine net is required in regions where the velocities are large, but to extend this into regions of low velocity would entail great labour and would bring no compensating advantage. What is wanted is a technique involving specially fine nets *in regions arbitrarily delimited by the computer*. This (a new development) we now explain.

Example 1

15. It may be illustrated by a problem which entails only one boundary condition—the torsion problem for a rectangular bar having two symmetrical keyways (figure 6). The 'stress-function' Ψ must satisfy the equation

$$\nabla^2\Psi+2 = 0 \tag{33}$$

at every point in the section, and it must have a constant value (which without loss of generality may be taken as zero) at every point on the boundary (cf. Part III, § 19). It is known that it will have, close to the keyways, gradients which are large compared with its gradients elsewhere; so a coarse net will suffice over most of the section, compared with the net which is appropriate there. On account of symmetry, only one-quarter of the section needs to be reproduced in figure 6.

In that diagram, outside of the region *abc* the mesh-side *a* is one-twelfth the length of a shorter side of the rectangle, or one-sixteenth the length of a longer side. (The rectangle has sides in the ratio 4 : 3.) Within *abc*, by insertion of diagonals, the meshes have been successively reduced in size (and rotated through 45°) until, within the

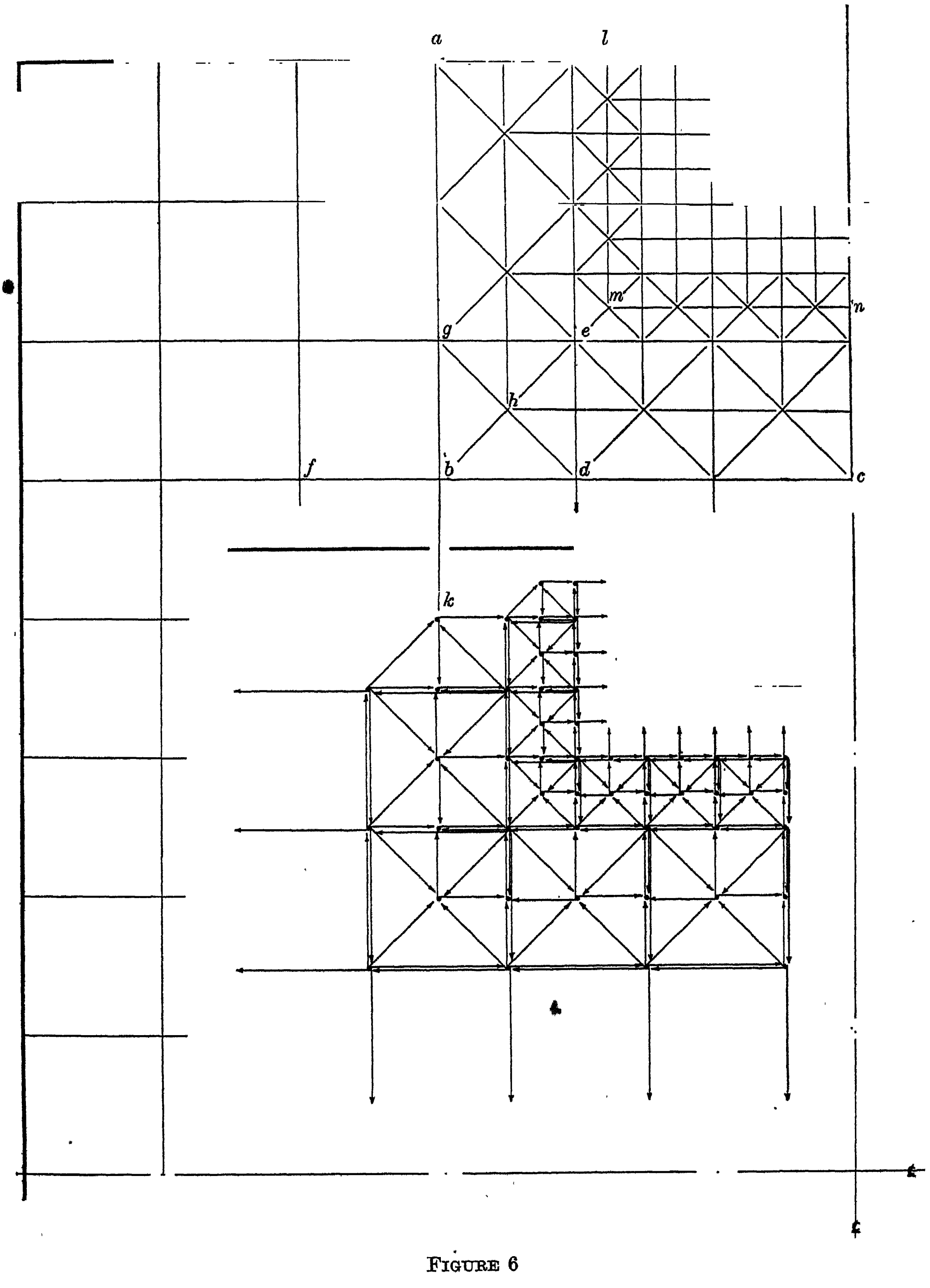


FIGURE 6

region *lmn*, the mesh-side is only one-quarter of what it is outside of *abc*, and equal to the radius at each inner corner of the keyway. Near to such corners, therefore, Ψ-values are available at nodal points sufficiently close to permit an accurate assessment of the Ψ-gradients (i.e. shear-stress components).

16. In 'liquidating residuals' on a net thus 'graded', care must be taken to ensure that correct 'relaxation patterns' are employed. Thus the residual force at h (figure 6) depends upon the Ψ-values at h, g, b, d, e, but the residual force at b depends upon the Ψ-values at b, g, f, k, d, which are nodes of a coarser net; consequently Ψ_b influences $\mathbf{F}_h$ but Ψ_h does not influence $\mathbf{F}_b$, and the notion (§ 12) of a force transmitted from h to b by the 'string' hb, as the result of a 'displacement' Ψ_h, must be discarded.

The 'relaxation pattern' for any nodal point O is a means of showing graphically at what points 'residuals' are affected by a displacement of O. From what has been said it follows that h is included in the pattern appropriate to b, but that b is not included in the pattern appropriate to h. In figure 6, to assist an understanding of the liquidation process, an inserted diagram shows 'patterns' drawn in accordance with obvious conventions. *These will not be needed by an experienced computer*, who will see at once the consequences of a given 'point-displacement'.

17. But the fact (§ 16) that a point-displacement will not necessarily entail transmission of force by every radiating 'string' means that apparently the sum of the residual forces will not be invariant for every operation, as it is when uniform nets are employed. The point e, for example, is 'fictitious' in relation to the coarsest net of figure 6, and force transmitted to it as a consequence of displacement of g or d is not brought into account because $\mathbf{F}_e$ (as now defined) does not depend on Ψ_g or on Ψ_d. This circumstance entails no difficulty, and the solution given in figure 7 was in fact obtained (by G.V.) easily and more quickly than it would have been if the finest net had extended to every part of the computational field. Ψ-contours show the direction and (by the closeness of their spacing) the magnitude of the shear stress at every point.

USE OF 'GRADED NETS': (2) WHEN BOUNDARY GRADIENTS ARE SPECIFIED

18. For nets having uniform mesh-side a in every part, a technique applicable to plane-potential problems of the 'second class' (i.e. problems in which the boundary condition imposes values on the normal gradient) was propounded in Part VIII, §§ 11–13. This in effect replaced the specified membrane by one having a slightly larger polygonal boundary, the sides consisting either of strings of a chosen net or of lines which bisect its meshes (diagonals when these are square, medians when they are triangular). Clearly the substituted membrane should be no larger than is necessary, so here too advantage may be expected to accrue from a use of 'graded nets' having specially fine meshes near the specified boundary.*

---

* The same will be true of 'Method 2' of Part VIII, §§ 20–23, which is applicable to 'mixed' boundary conditions.

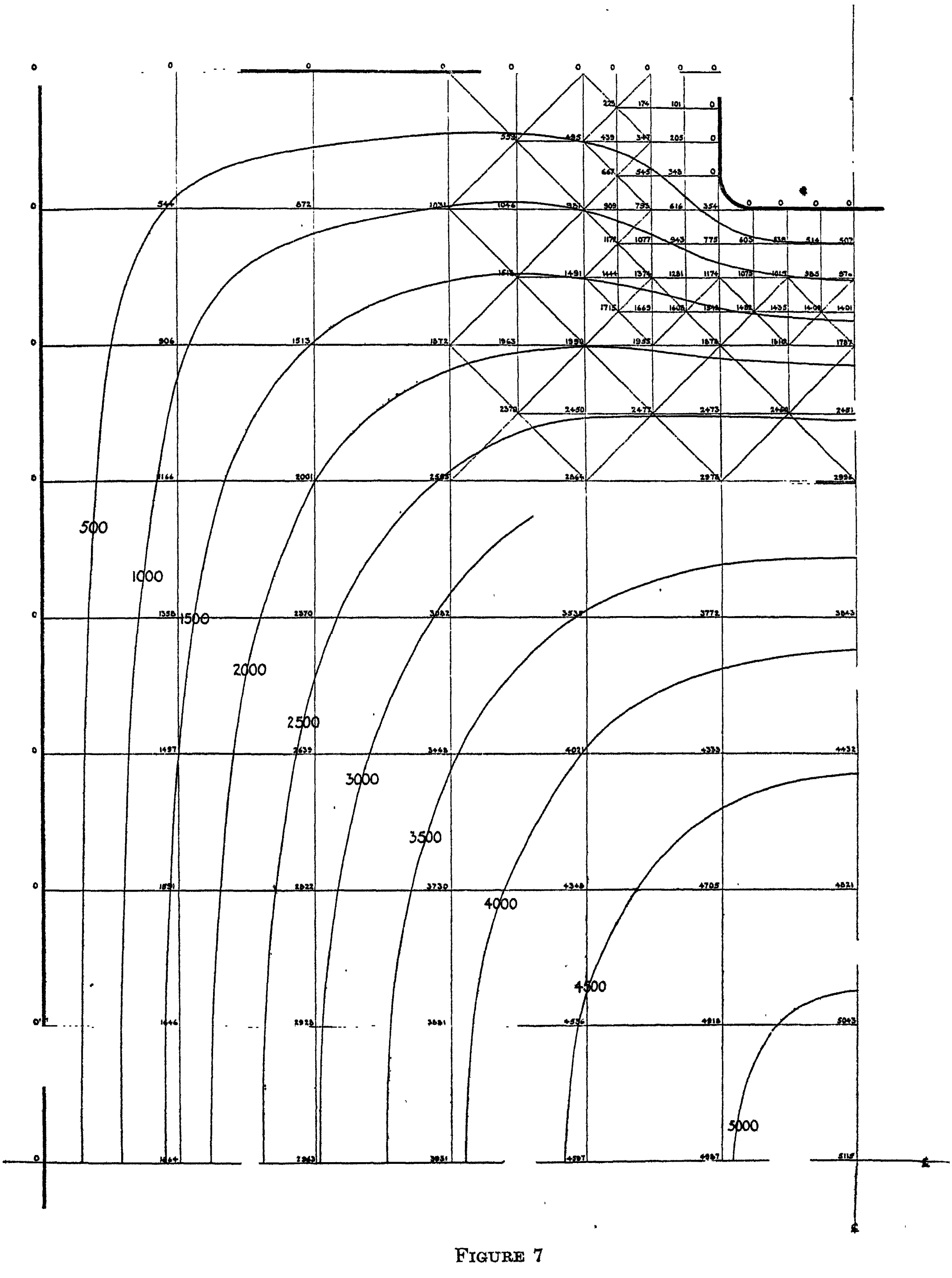


FIGURE 7

The technique will now be extended to 'quasi-plane-harmonic' problems where (§ 7) the membrane tension T is non-uniform; the objective being to solve problems of the 'second class' (normal gradients specified) with the highest accuracy that is attainable without excessive labour. The extension is based on a new interpretation of the standard formulae (21), § 8, and (26), § 11; and it will be found that this suggests a modification, to improve its accuracy, in the treatment of edge-loading which was propounded in Part VIII. The effect is not important when the net is fine.

ALTERNATIVE DERIVATION OF THE STANDARD FORMULAE. EXTENSION TO ANY POLYGONAL AREA

19. The new interpretation of (21) and (26) is based on the exact governing equation (1). Multiplied through by T, and integrated in respect of any area in which it applies, this yields a condition of equilibrium for the corresponding element of loaded membrane. With a use of Green's transformation, and substituting from (25), we obtain

$$\oint T\frac{\partial\psi}{\partial\nu}\,ds+\iint \mathsf{T}Z\,dx\,dy=0, \tag{34}$$

in which (§ 7) $\mathsf{T}Z$ denotes the intensity of transverse pressure and (§ 10) T denotes the membrane tension at any point. The surface integral measures the resultant of the pressure, the contour integral measures the force exerted by the membrane tension.

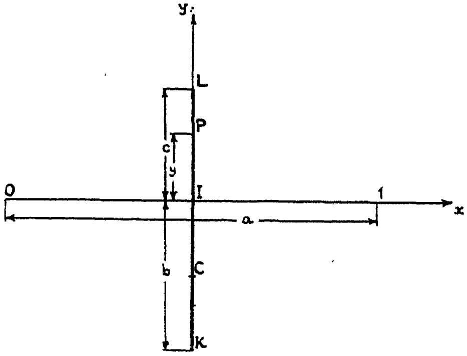


FIGURE 8

Estimation of the surface integral presents no difficulty, and the contour integral also can be evaluated for a region bounded by rectilinear elements. In figure 8, let KL be an element of this kind and 01 a line of length a which cuts it perpendicularly at I and which is itself bisected at that point. Then, in the notation of the diagram,

$$\left.\begin{aligned}\chi_P &= \chi_{\mathrm{I}}+y\left(\frac{\partial\chi}{\partial y}\right)_{\mathrm{I}}+\tfrac{1}{2}y^2\left(\frac{\partial^2\chi}{\partial y^2}\right)_{\mathrm{I}}+\dots,\\ \left(\frac{\partial\psi}{\partial x}\right)_P &= \left(\frac{\partial\psi}{\partial x}\right)_{\mathrm{I}}+y\left(\frac{\partial^2\psi}{\partial x\,\partial y}\right)_{\mathrm{I}}+\tfrac{1}{2}y^2\left(\frac{\partial^3\psi}{\partial x\,\partial y^2}\right)_{\mathrm{I}}+\dots,\end{aligned}\right\} \tag{i}$$

and hence for the element KL, if Ox is the direction of the outward normal ν,

$$\int T\frac{\partial\psi}{\partial\nu}\,ds = \mathsf{T}\int_K^L \chi\frac{\partial\psi}{\partial x}\,dy$$

$$= \mathsf{T}\int_{-b}^{c}\left[\chi_{\mathrm{I}}\left(\frac{\partial\psi}{\partial x}\right)_{\mathrm{I}} + y\left\{\left(\frac{\partial\psi}{\partial x}\right)_{\mathrm{I}}\left(\frac{\partial\chi}{\partial y}\right)_{\mathrm{I}} + \chi_{\mathrm{I}}\left(\frac{\partial^2\psi}{\partial x\,\partial y}\right)_{\mathrm{I}}\right\} + (\text{terms in } y^2,\ y^3,\ \text{etc.})\right]dy,$$

$$= \mathsf{T}\left[(b+c)\,\chi_{\mathrm{I}}\left(\frac{\partial\psi}{\partial x}\right)_{\mathrm{I}} + \tfrac{1}{2}(c^2-b^2) \times\left\{\left(\frac{\partial\psi}{\partial x}\right)_{\mathrm{I}}\left(\frac{\partial\chi}{\partial y}\right)_{\mathrm{I}} + \chi_{\mathrm{I}}\left(\frac{\partial^2\psi}{\partial x\,\partial y}\right)_{\mathrm{I}}\right\}\right] \qquad \text{(ii)}$$

when terms of order b^3, b^4, ..., c^3, c^4, ..., etc., are neglected.

20. Three cases have practical interest. First, when $c = b$ in figure 8, then (ii) may be written as

$$\int_K^L T\frac{\partial\psi}{\partial\nu}\,ds = 2b\mathsf{T}\chi_{\mathrm{I}}\left(\frac{\partial\psi}{\partial x}\right)_{\mathrm{I}}, \qquad \text{(iii)}$$

with the same approximation as before; and since

$$\left.\begin{aligned}\psi_1-\psi_0 &= a\left(\frac{\partial\psi}{\partial x}\right)_{\mathrm{I}} + \tfrac{1}{24}a^3\left(\frac{\partial^3\psi}{\partial x^3}\right)_{\mathrm{I}} + \dots\text{etc.},\\ \chi_0+\chi_1 &= 2\left[\chi_{\mathrm{I}} + \frac{a^2}{8}\left(\frac{\partial^2\chi}{\partial x^2}\right)_{\mathrm{I}} + \dots\text{etc.}\right],\end{aligned}\right\} \qquad (35)$$

it follows that with neglect (in comparison with unity) of a^2 and b^2 and of their higher powers the right-hand side of (iii) may be replaced either by $2b\mathsf{T}\chi_{\mathrm{I}}(\psi_1-\psi_0)/a$ or by $b\mathsf{T}(\chi_0+\chi_1)\,(\psi_1-\psi_0)/a$. To the same approximation, the resultant of the membrane tension across KL

$$= \int_K^L T\,dy = \mathsf{T}\int_{-b}^{b}\chi_P\,dy = 2b\mathsf{T}\chi_{\mathrm{I}}, \quad \text{according to (i)},$$

$$= b\mathsf{T}(\chi_0+\chi_1), \quad \text{according to (35)};$$

consequently (iii) may be written in the form

$$\left.\begin{aligned}\int T\frac{\partial\psi}{\partial\nu}\,ds \text{ for the element } KL &= \mathbf{T}_1\frac{\psi_1-\psi_0}{a},\\ \text{where}\qquad \mathbf{T}_1 &= 2b\mathsf{T}\chi_{\mathrm{I}} = b\mathsf{T}(\chi_0+\chi_1)\end{aligned}\right\} \qquad (36)$$

and the line 01 has the direction of the *outward* normal.

Using this result we can formulate a condition of equilibrium for *any* polygonal element, thereby generalizing the results of § 11. The formula (36) confirms these in the case (contemplated in § 11) of a regular polygon having N sides of length $a\tan\pi/N$; for on substituting $a\tan\pi/N$ for $2b$ in (36) we have

$$\mathbf{T} = a\mathbf{T}\tan\frac{\pi}{N}\,\chi_{\mathrm{I}}, \tag{37}$$

agreeing with (27) when (as in § 11) the factor $\mathbf{T}\tan\pi/N$ is equated to unity.

21. The second case having practical interest is that in which $c \doteq 0$ in figure 8 *and the line* 01 *is part of an unloaded boundary*. On that understanding $\partial\psi/\partial y = 0$ everywhere along 01, and accordingly $(\partial^2\psi/\partial x\,\partial y)_{\mathrm{I}}$ in (ii) is zero; consequently if $c = 0$ then *with the same approximation as before* we have in place of (ii)

$$\begin{aligned}
\int_K^L \frac{\partial\psi}{\partial\nu}\,ds &= b\mathbf{T}\left(\frac{\partial\psi}{\partial x}\right)_{\mathrm{I}}\left\{\chi_{\mathrm{I}} - \tfrac{1}{2}b\left(\frac{\partial\chi}{\partial y}\right)_{\mathrm{I}}\right\},\\
&= b\mathbf{T}\chi_C\left(\frac{\partial\psi}{\partial x}\right)_{\mathrm{I}}, \quad \text{according to (i), } C \text{ denoting the middle point of } IK \text{ in figure 8,}\\
&= \mathbf{T}_{\frac{1}{2}}\frac{\psi_1-\psi_0}{a}, \quad \text{according to (35),}\\
\text{where}\quad \mathbf{T}_{\frac{1}{2}} &= b\mathbf{T}\chi_C.
\end{aligned} \tag{38}$$

The interpretation is the same as in § 20, except that now $\mathbf{T}_{\frac{1}{2}}$, in place of $\mathbf{T}_1$, is the tension to be associated with the string 01. When $\chi = 1$ everywhere, $\mathbf{T}_{\frac{1}{2}}$ according to (38) $= \frac{1}{2}\mathbf{T}_1$ according to (36), so the result given confirms the 'half-tension' rule of Part VIII, § 12, in respect of a string which coincides with an element of a polygonal boundary. But our examination has not only generalized that rule for use in problems when χ is variable, it has also shown that the rule has less than the standard order of approximation if the element of boundary is 'loaded', because in that event (since $(\partial^2\psi/\partial x\,\partial y)_{\mathrm{I}}$ in (ii) no longer vanishes) a term of order b has been neglected (in comparison with unity) in deriving (38).

22. The third case having practical importance is that in which $\partial\psi/\partial\nu \doteq 0$ at all points in KL (figure 8). Then $\int_K^L T\frac{\partial\psi}{\partial\nu}\,ds$ is zero, so no tension has to be given to the string 01 which is associated with the element KL. In other words, *no string need be associated with an unloaded element of boundary.*

APPLICATION TO THE PROBLEM OF SPECIFIED BOUNDARY GRADIENTS

23. To see the application of these results, assume (in the first place) that A (figure 9) is a polygonal boundary lying wholly outside of the specified boundary B. As in Part VIII, § 3, it may be said that specified values of $\partial\psi/\partial\nu$ are specified values of $T(\partial\psi/\partial\nu)$, so the problem presented is to compute the effects of a specified transverse loading applied to B. *This loading we apply to B regarded as a line traced on the larger polygonal membrane, and we assume the outer boundary A to be unloaded.*

Clearly the solution of the problem as thus modified will differ from the wanted solution (for a membrane having the shape of B) only in so far as constraint is exerted by the membrane lying between B and A.

This added membrane (shaded in figure 9) may be termed the 'selvedge'. It must (in general) exercise *some* constraint, because variation of ψ along B will entail curvature in the direction of the normal to B and so the normal gradient, though zero at all points of A, will not be zero on B; but clearly the constraint will decrease with the selvedge width, so the purpose must be to make this small. To this end the results of §§ 21 and 22 are utilized: Because A's edge is unloaded no string of the relaxation net need cross it, and where convenient A may be drawn to pass through two adjacent nodal points (i.e. to coincide with a string) provided that with that string is associated a 'half-tension' $T_{\frac{1}{2}}$ as given by (38). The extent of the relaxation net is determined by the consideration that strings must cross the loaded boundary B in order that the specified loading may be transmitted to nodal points.

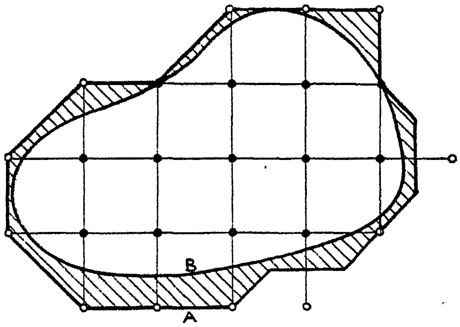
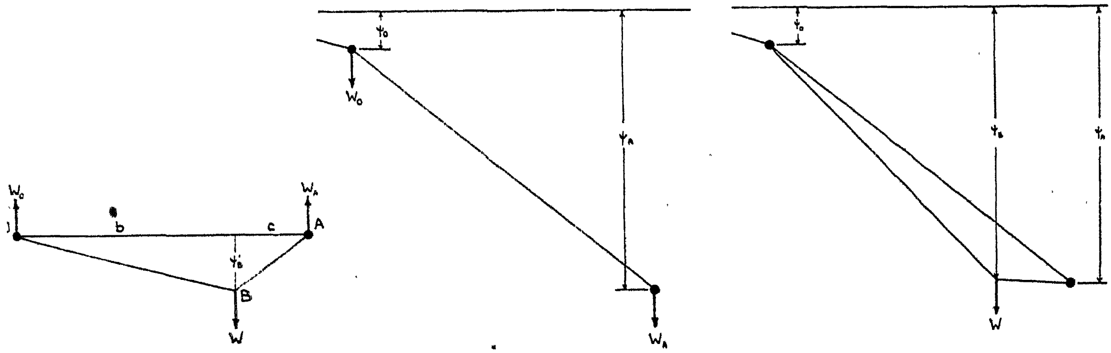


FIGURE 9



(*a*) (*b*) (*c*)

FIGURE 10

24. It is here that we depart (in plane-harmonic problems) from the treatment which was given in Part VIII. There (§ 7) it was remarked that if distributed loading is to be brought to bear on a net in such a way that all strings remain straight when

the net is deflected, some mechanism must be envisaged for effecting the distribution: clearly the same is true of load which instead of being distributed over a surface is applied along a line as in § 23, and here an appropriate mechanism will consist of rigid rods conforming with the line of loading (short elements of which may be treated as straight) and transmitting the applied forces to their end-points on the strings. If as before the condition is imposed that all strings of the deflected net remain straight, other rigid rods will be needed to transmit these (concentrated) forces to nodal points; but then the deduced deflexion will have no discontinuity at the line of loading, so a wanted feature will be lost. *Accordingly we now dispense with this last requirement as regards strings which are cut by* B (§ 23): instead, we assume that these by deflecting transmit their loads to their (nodal) end-points.

25. Let OBA (figure 10) be a string loaded at B with a concentrated force W, and assume in the first place that its ends O, A (which are nodal points) do not deflect (figure 10*a*). Then, in the notation of the diagram, forces

$$W_O = \frac{c}{a}W, \quad W_A = \frac{b}{a}W \quad (b+c = a)$$

will be required, for equilibrium, at O and A respectively; and the deflexion of B will be

$$bcW/a\mathrm{T} = \psi'_B, \text{ say,} \tag{i}$$

T denoting the tension in the string OA. Now let forces equal and opposite to W_O, W_A act at O and A, and let ψ_0, ψ_A (figure 10*b*) be the displacements (determinable by relaxation methods) which result from these and corresponding forces at other points. The consequent deflexion of B will be

$$\psi''_B = (c\psi_0 + b\psi_A)/a, \tag{ii}$$

and if, finally, the solutions (i) and (ii) are superposed *so as to neutralize the forces at* O *and* A, the resultant displacement of B will be (figure 10*c*)

$$\psi_B = \psi'_B + \psi''_B = \frac{c}{a}\left(\psi_0 + \frac{bW}{\mathrm{T}}\right) + \frac{b}{a}\psi_A. \tag{39}$$

The displacements of internal nodes (typified by ψ_0) have been found already, and using (39) we can compute the boundary displacements, thus completing the solution.

Because the relation (34), § 19, is derived from the exact equation (1) and so must hold in respect of the whole of the given membrane (i.e. of the whole area within the line of loading B, § 23), by assuming no force to act upon the selvedge (between B and A) we shall be left with forces which sum to zero: consequently the relaxational problem will be tractable, as was argued in Part VIII, § 11. The treatment here described of loaded strings differs from what was described in § 12 of that paper, and may be expected to result in greater accuracy.

Example 2

26. Figure 11 presents a simple problem which will serve to exemplify these methods and (since the exact solution is known) to test their accuracy. *The membrane (and hence the net) is plane in the exact solution, whatever be the distribution of* χ: therefore for simplicity it has been assumed that $\chi = 1$ everywhere, so that the governing equation is plane-harmonic.

Exact values of the displacement (ψ) are recorded in figure 11 at nodal points of one of the (square-mesh) nets that were used in computation. The corresponding gradients are calculable, so the problem could be treated on the understanding that normal gradients of ψ are specified along the edges AB, BC, CD of the trapezoidal strip shown in thick line, values of ψ are specified along the base AD.

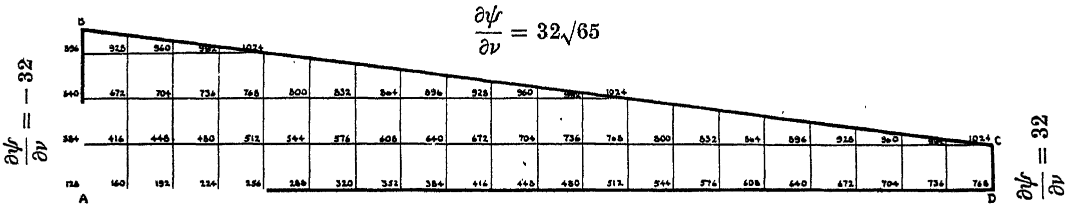


FIGURE 11. In calculating gradients 1 mesh-length has been taken as unity.

27. Figure 12*a*, *b* exhibit the relaxational solution of this problem (by G. V.) on two successive nets. Internal nodes are shown by black dots, strings which cross the loaded boundary are drawn to fictitious nodes distinguished by open circles, and the extent of the postulated 'selvedge' is indicated, i.e. the shape of the unloaded outer boundary (A, § 23).

The correct value of ψ is 1024 at all points of BC, and along BA and CD ψ should vary linearly: even in figure 12*a* the computed values nowhere have an error exceeding 1·6 %, and in figure 12*b* the greatest error has fallen to 0·9 %. The correct value of the vertical gradient ($\partial\psi/\partial y$) is 256 at all points on BC: computed values of this quantity are plotted in the upper part of figure 12*a*, and the errors are seen to be greatest at B and C and where BC is cut by horizontal strings. This is because eight vertical strings meet BC for every one horizontal string: in the correct solution the horizontal gradient ($\partial\psi/\partial x$) has a constant value along BC and so accounts for a constant part of the uniform loading, but in the net this loading has to be resisted by vertical strings alone, except at three points where a horizontal string contributes. In figure 12*b* a similar plotting shows errors having a like explanation but slightly reduced in magnitude on account of the finer mesh.

Example 3

28. In Example 2 (consistently with § 23) the polygonal boundary A was made either to coincide with or to lie *outside* the specified boundary B. But our aim in making the selvedge width small is to reduce to a minimum the additional constraint which it imposes, and clearly there is no objection to a 'negative

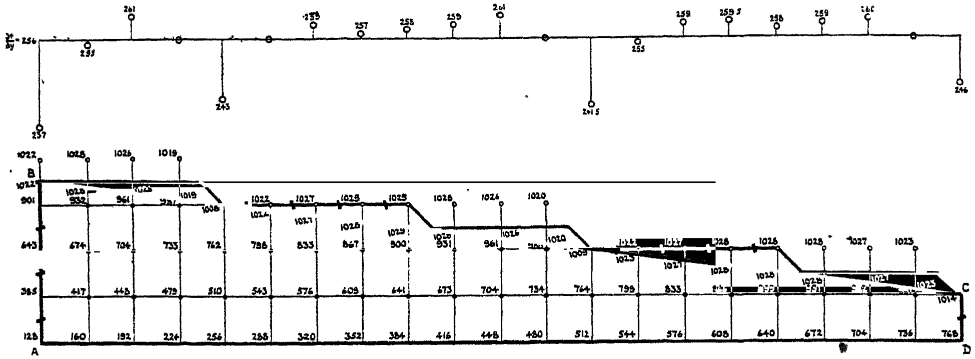


Figure 12*a*. ‘Half-tension strings’ are indicated by ticks.

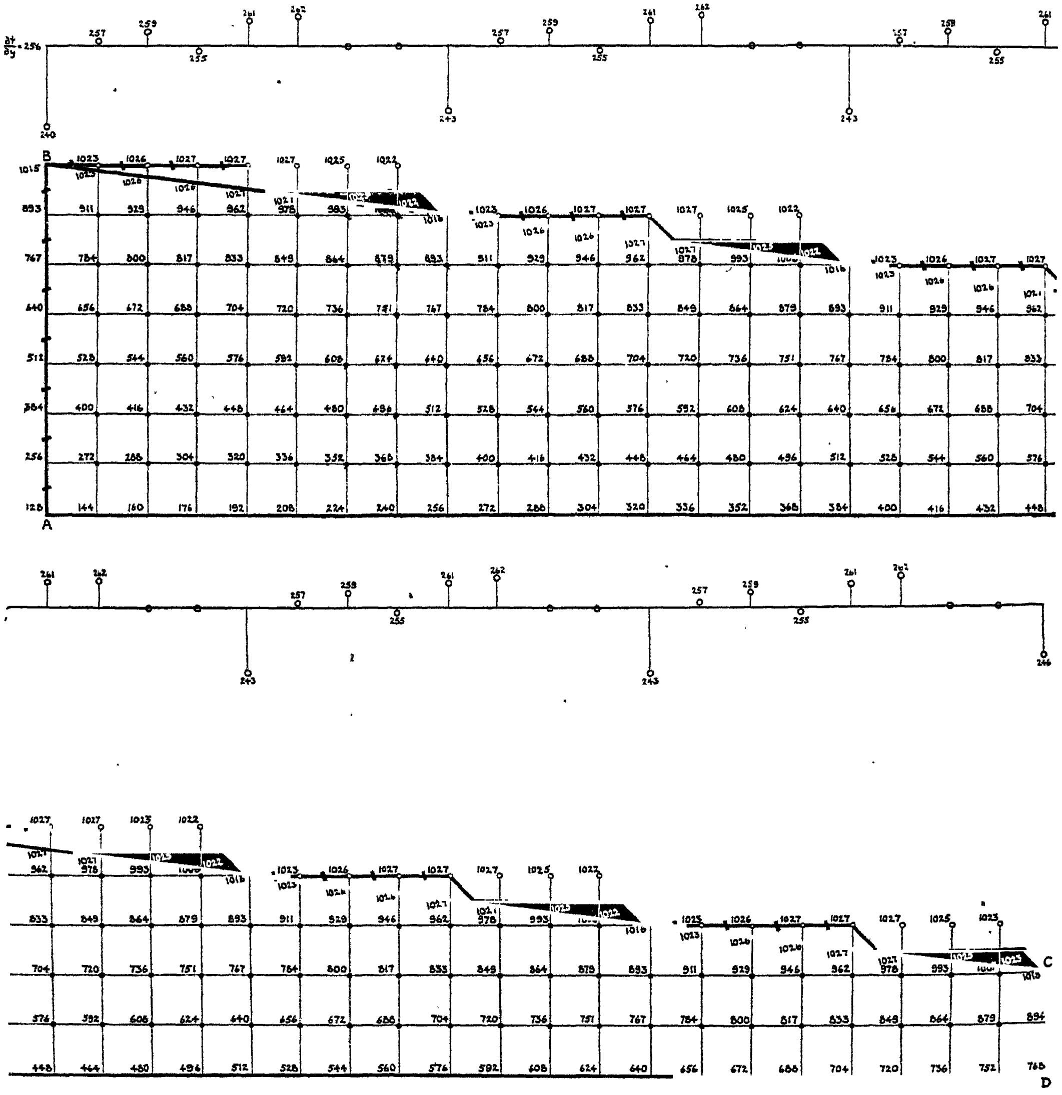


Figure 12*b*. In calculating gradients 2 mesh-lengths have been taken as unity.

selvedge' (i.e. to letting A lie *inside* B), provided that the consequent *reduction* of constraint is small. In Example 3 (also a plane-harmonic problem) the chosen selvedge was in places positive but in other places negative, as shown in figure 13. Intuitively it is evident that the projecting 'tabs' (of 'positive selvedge') will exercise little constraint; for they embrace only short lengths of the boundary, and their outer edges are unloaded.

This problem was treated in Part VIII, §§ 14–15, with results which were recorded in figures 9 of that paper. The specified normal gradients were chosen so as to make

$$w = \tan^{-1}(y/x) - \tfrac{1}{4}\pi \qquad (40)$$

an exact solution (cf. Part VIII, equations (18), § 14, and (22), § 15). The errors of the relaxational solution were small, so cannot be much reduced by modifying the technique of computation in the manner suggested here: therefore to show the improvement we have thought it best to record the 'residuals' which are left, according to the original and to the amended technique, *when the exact solution* (40) *is assumed for the displacements*. (An exact technique would yield zero residuals, and the more exact technique is that in which the residuals have the smaller values. Here, both techniques leave some residuals in all parts of the field of computation, because (as was remarked in Part VIII, § 16) the wanted function (40) does not exactly satisfy the relation

$$\Sigma_{a,6}(w) - 6w_0 = 0,$$

which is the finite-difference approximation to the governing equation.)

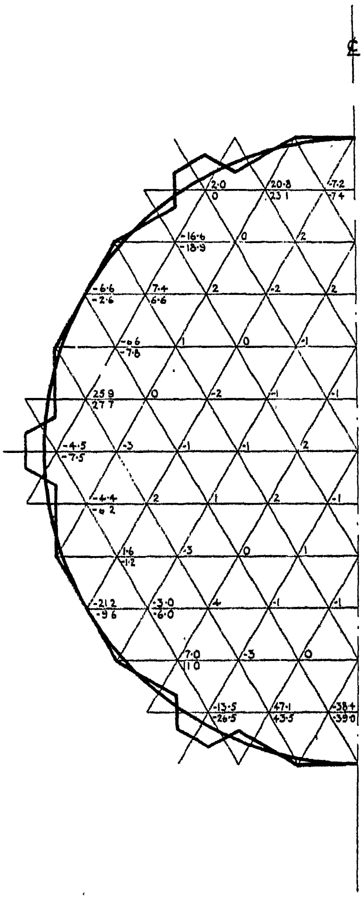


FIGURE 13

29. In figure 13, at internal nodes which adjoin the circular boundary, numerals give the residuals computed in accordance with this paper (above) and with Part VIII (below). For this purpose, at internal nodes w was computed from (40), and at the fictitious nodes it was given their values required (according to each technique in turn) to make the residuals zero. Thereby residuals were transferred from the fictitious to the internal nodes, and the resulting values are recorded. Comparison reveals that a small but definite advantage is possessed by the modified technique: the algebraic sum of the compared residuals is $-20{\cdot}8$ according to Part VIII, $-10{\cdot}2$ according to the present paper; and their 'root-mean-square'

value is 19·2 according to Part VIII, 18·8 according to the present paper.* (The total of the applied load on the half circle was 2976.)

Example 4

30. Our last example is concerned with thermal conduction in a system having axial symmetry. Its governing equation is (15) of § 5, so it is a quasi-plane-potential problem.

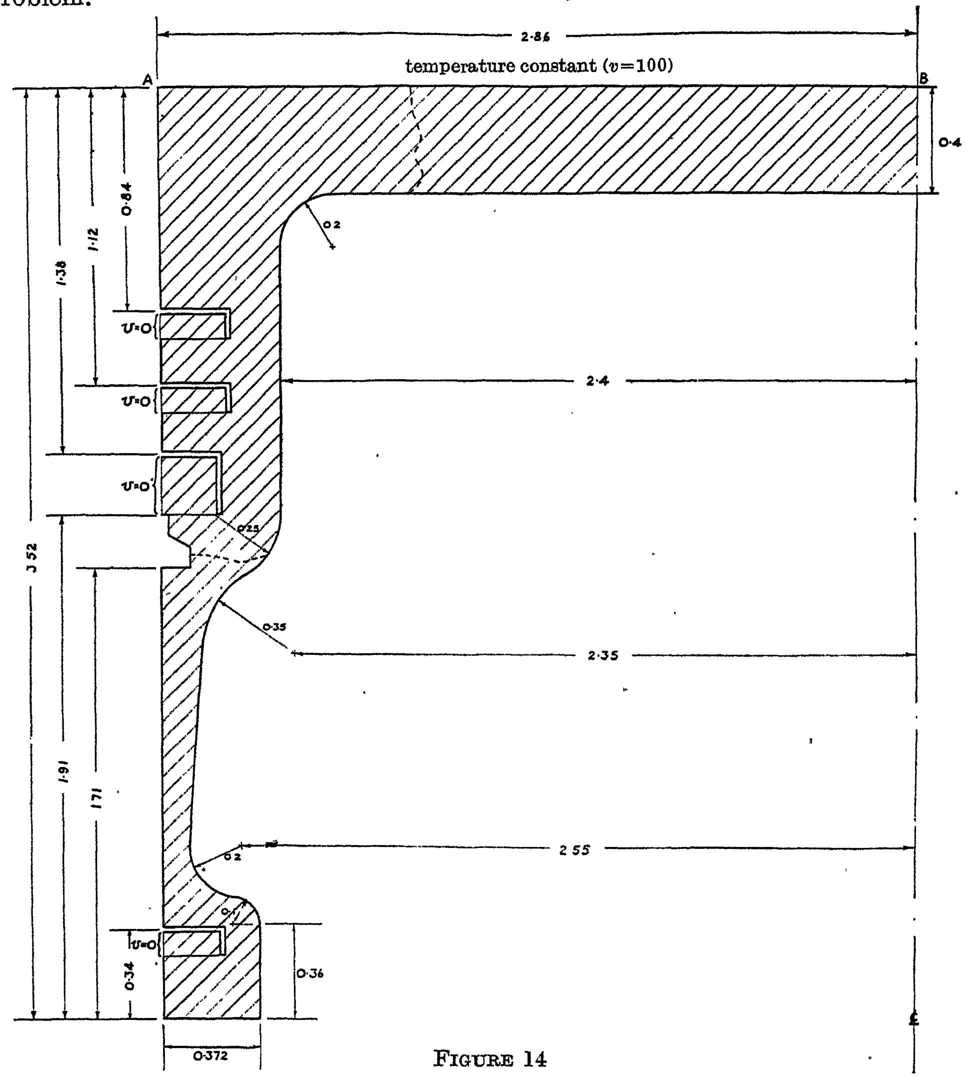


FIGURE 14

The piston of an internal combustion engine receives heat from the products of combustion through its upper surface and transmits this mainly to the cylinder walls, which are cooled: some heat is transmitted through its lower surface to the

---

* The algebraic sum is the better criterion, because positive and negative values tend in the relaxation process to meet and cancel.

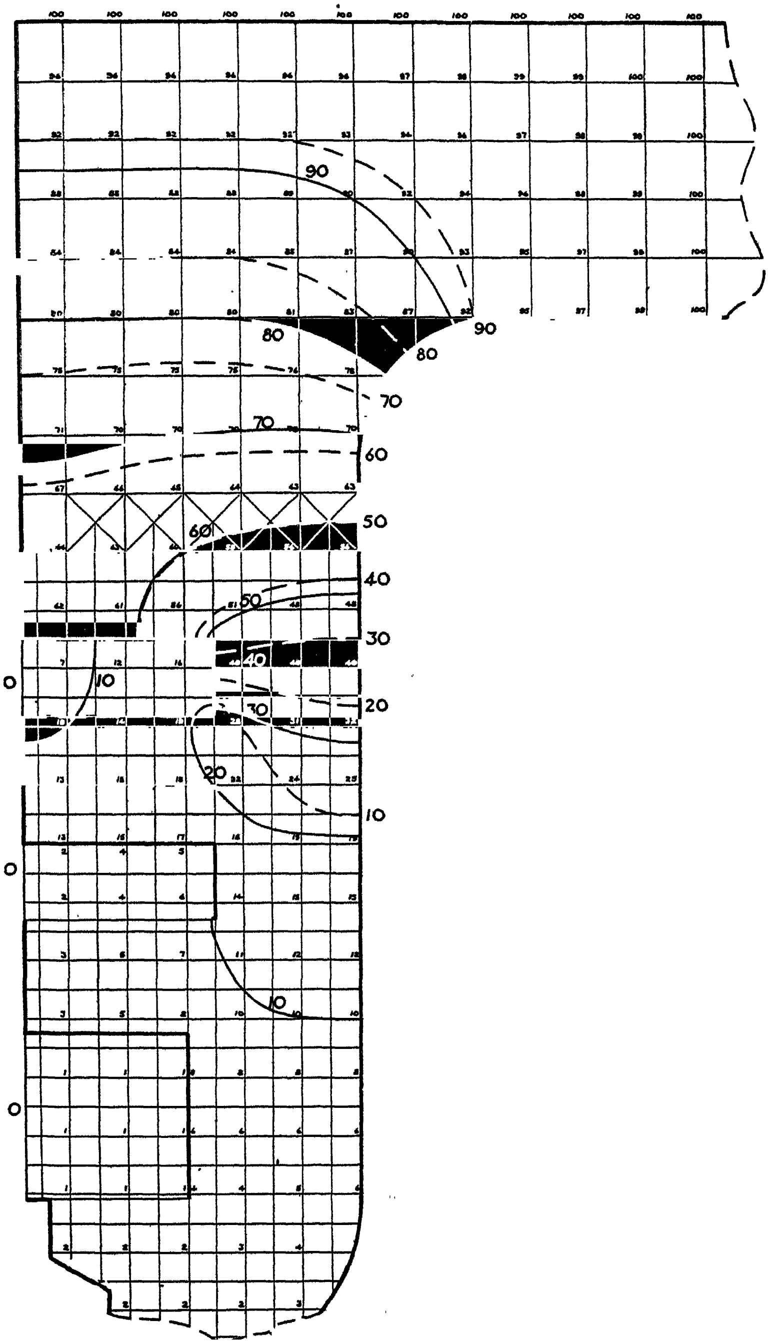


FIGURE 15

contents of the crank-case, but this is relatively small and here it will be neglected. Similarly of the heat transmitted to the cylinder walls, because oil has low conductivity almost all goes through those surfaces which are brought most nearly into contact: these are (1) those surfaces of each piston ring and groove which are pressed together by reason of the (frictional and inertial) resistance to the piston's motion, (2) the outer surface of each piston ring and that part of the cylinder wall against which it presses.

Thus, by simplification of the physical conditions, the idealized computational problem shown in figure 14 is obtained. *No* heat is assumed to pass through any of the surfaces (other than the heated top surface) which are drawn in thick line, and *no* insulating effect is attributed to the *thin* oil films which separate the piston rings from their grooves and from the cylinder walls (so that in effect continuous metal is assumed to occupy the space shown by continuous shading). On the upper surface of the piston (AB, figure 14) the temperature is assumed to have a constant $v_1 = 100$; and the cylinder wall, with those surfaces of the piston rings which touch it, are assumed to have another constant temperature $v_0 = 0$. The problem is to compute v for points within the metal of the piston, and thence to deduce the associated heat-flow. Clearly the isothermals will be 'crowded' in the neighbourhood of the piston rings, so a use of graded nets is indicated.

31. Computation (by D. N. de G. A.) confirmed the expected 'crowding' and revealed that practically no heat is transmitted, in these circumstances, through the lowest of the piston rings. Consequently only a portion of the computation field (indicated by broken lines in figure 14) is represented in figure 15. There the grading of the relaxation net is shown, accepted values of v are recorded, and isothermal lines are drawn for intervals of 10 units.

It was suggested to us by Mr J. F. Alcock that in future special alloys may be used for piston rings on account of the dominating part which these play in heat transmission, and that therefore it would be useful to investigate the consequences of improving their thermal conductivity. Accordingly it was decided to examine what would be the temperature distribution if the rings were *infinitely* conducting (so that $v = 0$ on the surface of contact of each ring and groove). Broken-line contours in figure 15 exhibit the isothermals in this limiting case. They call for no remark in this paper, excepting as they indicate that relaxation methods will deduce the consequences of *any* assumption that may be made.

CONCLUSION

32. When $\chi = 1$ everywhere in the computational field, equation (1), § 1, reduces to the 'plane-harmonic' (Poisson) equation which has been treated in previous papers. This paper examines the more general case in which χ is a specified function of x and y.

For orthodox methods the general case is far more difficult, but negligible modifications are required in the relaxational technique developed in this series. The 'net analogue' holds as before, and *it is necessary only to associate with each 'string' a tension* (T) *equal to the value of χ at its middle point* (cf. § 12), except in circumstances discussed in § 21, where equation (38) states a new generalization of the 'half-tension rule' of Part VIII, § 12. The investigation has yielded a new interpretation of the basic finite-difference approximation, and this in turn has led to slight modifications of the technique propounded in Part VIII for problems of 'Class 2' (normal gradients specified).

Thus *from a relaxational standpoint* the 'quasi-plane-potential' equation differs negligibly from the plane-harmonic, and what is new in the present paper applies with equal force to both equations. For that reason plane-harmonic problems were chosen for three of the four examples, as being more familiar. The thermal problem solved in figure 15 will suffice to show the power of the relaxational technique.

REFERENCES

Relaxation methods applied to engineering problems:

Part I:	Bradfield, K. N. E. & Southwell, R. V. 1937 *Proc. Roy. Soc.* A, **161**, 155–181.
Part III:	Christopherson, D. G. & Southwell, R. V. 1938 *Proc. Roy. Soc.* A, **168**, 317–350.
Part V:	Gandy, R. W. G. & Southwell, R. V. 1940 *Phil. Trans.* A, **238**, 453–475.
Part VI:	Pellew, A. & Southwell, R. V. 1940 *Proc. Roy. Soc.* A, **175**, 262–290.
Part VII:	Shaw, F. S. & Southwell, R. V. 1941 *Proc. Roy. Soc.* A, **178**, 1–17.
Part VIII:	Southwell, R. V. & Vaisey, G. 1943 *Proc. Roy. Soc.* A, **182**, 129–151.
Part IX:	Green, J. R. & Southwell, R. V. 1944 *Phil. Trans.* A, **239**, 367–386.

Christopherson, D. G. 1941 *Proc. Instn Mech. Engrs*, **146**, 125–135.
Southwell, R. V. 1935 *Proc. Roy. Soc.* A, **151**, 56–95, and **153**, 41–76.
Southwell, R. V. 1942 *Proc. Roy. Soc.* A, **180**, 367–396.

Unitary representations of the Lorentz group

By P. A. M. Dirac, F.R.S., *St John's College, Cambridge*

(*Received* 31 *May* 1944)

Certain quantities are introduced which are like tensors in space-time with an infinite enumerable number of components and with an invariant positive definite quadratic form for their squared length. Some of the main properties of these quantities are dealt with, and some applications to quantum mechanics are pointed out.

1. Introduction

Given any group, an important mathematical problem is to get a matrix representation of it, which means to make each element of the group correspond to a matrix in such a way that the matrix corresponding to the product of two elements is the product of the matrices corresponding to the factors. The matrices may be looked upon as linear transformations of the co-ordinates of a vector and then each element of the group corresponds to a linear transformation of a field of vectors. Of special interest are the *unitary* representations, in which the linear transformations leave invariant a positive definite quadratic form in the co-ordinates of a vector.

The Lorentz group is the group of linear transformations of four real variables $\xi_0, \xi_1, \xi_2, \xi_3$, such that $\xi_0^2-\xi_1^2-\xi_2^2-\xi_3^2$ is invariant. The finite representations of this group, i.e. those whose matrices have a finite number of rows and columns, are all well known, and are dealt with by the usual tensor analysis and its extension spinor analysis. None of them is unitary. The group has also some infinite representations which are unitary. These do not seem to have been studied much, in spite of their possible importance for physical applications.

The present paper gives a new method of attack on these representations, which was suggested by Fock's quantum theory of the harmonic oscillator. It leads to a new kind of tensor quantity in space-time, with an infinite number of components and a positive definite expression for its squared length.

2. Three-dimensional theory

This section will be devoted to some preliminary work applying to the rotation group of three-dimensional Euclidean space. Take an ascending power series

$$a_0+a_1\xi_1+a_2\xi_1^2+a_3\xi_1^3+\dots \tag{1}$$

in a real variable ξ_1 with real coefficients a_r. Consider these coefficients to be the co-ordinates of a vector in a certain space of an infinite number of dimensions, and define the squared length of the vector to be

$$\Sigma_0^\infty r!\,a_r^2. \tag{2}$$

The series (2) must converge for the vector to be a finite one.

Take two more similar power series $\Sigma_0^\infty b_s\xi_2^s$ and $\Sigma_0^\infty c_t\xi_3^t$ in the real variables ξ_2 and ξ_3 and consider their coefficients to be the co-ordinates of vectors in two more vector spaces, with squared lengths defined by the corresponding formula to (2). Now multiply the three vector spaces together. A general vector in the product space will be a sum of products of vectors of the three original vector spaces, and its co-ordinates A_{rst} can be represented as the coefficients in a power series

$$P = \Sigma_0^\infty A_{rst}\xi_1^r\xi_2^s\xi_3^t \tag{3}$$

in the three variables ξ_1, ξ_2, ξ_3. The squared length of such a vector is

$$\Sigma_0^\infty r!\,s!\,t!\,A^2_{rst}, \tag{4}$$

and two vectors with co-ordinates A_{rst} and B_{rst} have a scalar product

$$\Sigma_0^\infty r!\,s!\,t!\,A_{rst}B_{rst}. \tag{5}$$

If the variables ξ_1, ξ_2, ξ_3 are subjected to a linear transformation, going over into ξ_1', ξ_2', ξ_3', say, the power series (3) will go over into a power series in ξ_1', ξ_2', ξ_3',

$$P = \Sigma A'_{rst}\xi_1'^r\xi_2'^s\xi_3'^t,$$

in which the coefficients A' are linear functions of the previous coefficients A. Thus each linear transformation of the ξ's generates a linear transformation of the coefficients A.

The theorem will now be proved: *A linear transformation of ξ_1, ξ_2, ξ_3 which leaves $\xi_1^2+\xi_2^2+\xi_3^2$ invariant generates a linear transformation of the coefficients A which leaves the squared length* (4) *invariant.* Consider first the infinitesimal transformation

$$\xi_1 = \xi_1'+\epsilon\xi_2', \quad \xi_2 = \xi_2'-\epsilon\xi_1', \quad \xi_3 = \xi_3', \tag{6}$$

which leaves $\xi_1^2+\xi_2^2+\xi_3^2$ invariant, ϵ being a small quantity whose square is negligible. Substituting into (3), one gets

$$P = \Sigma A_{rst}(\xi_1'^r\xi_2'^s+r\epsilon\xi_1'^{r-1}\xi_2'^{s+1}-s\epsilon\xi_1'^{r+1}\xi_2'^{s-1})\,\xi_3'^t.$$

Hence $$A'_{rst} = A_{rst}+(r+1)\,\epsilon A_{r+1,s-1,t}-(s+1)\,\epsilon A_{r-1,s+1,t},$$

in which A_{rst} with a negative suffix is counted as zero. Thus

$$\Sigma r!\,s!\,t!\,A'^2_{rst} = \Sigma r!\,s!\,t!\,[A^2_{rst}+2(r+1)\,\epsilon A_{rst}A_{r+1,s-1,t}-2(s+1)\,\epsilon A_{r-1,s+1,t}A_{rst}].$$

The last two terms in the [ ] here cancel, as may be seen by substituting $r-1$ for r in the former and $s-1$ for s in the latter, and hence the squared length (4) is invariant for the transformation (6). Any linear transformation of ξ_1, ξ_2, ξ_3 which leaves $\xi_1^2+\xi_2^2+\xi_3^2$ invariant can be built up from the infinitesimal transformation (6) and similar infinitesimal transformations with ξ_1, ξ_2 and ξ_3 permuted, together with possibly a reflexion $\xi_1 = -\xi_1'$, $\xi_2 = \xi_2'$, $\xi_3 = \xi_3'$, which obviously leaves the squared length (4) invariant, and hence the theorem is proved.

The group of transformations of the ξ's which leave $\xi_1^2+\xi_2^2+\xi_3^2$ invariant is the rotation group in three-dimensional Euclidean space, so the transformations of the coefficients A provide a representation of this rotation group. One may restrict the function P to be homogeneous, of degree u say, and then the representation is a finite one. The coefficients A then form the components of a symmetrical tensor of rank u, the connexion with the usual tensor notation being effected by taking A_{rst} to be $u!/r!\,s!\,t!$ times the usual tensor component with the suffix 1 occurring r times, 2 occurring s times and 3 occurring t times, as may be seen from the invariance of expression (3) with (ξ_1, ξ_2, ξ_3) transforming like a vector.

One can make a straightforward generalization of the foregoing theory by introducing other triplets of variables, η_1, η_2, η_3 and ζ_1, ζ_2, ζ_3 say, which transform together with ξ_1, ξ_2, ξ_3, and setting up a power series in all the variables. The transformations of the coefficients of these more general power series will provide further representations of the three-dimensional rotation group. If such a more general power series is restricted to be homogeneous, its coefficients will form the components of an unsymmetrical tensor.

3. FOUR-DIMENSIONAL THEORY

Take a descending power series

$$k_0/\xi_0 + k_1/\xi_0^2 + k_2/\xi_0^3 + k_3/\xi_0^4 + \dots \tag{7}$$

in a real variable ξ_0 with real coefficients k_n. Consider these coefficients to be the co-ordinates of a vector in a certain space of an infinite number of dimensions, and define the squared length of the vector to be

$$\Sigma_0^\infty k_n^2/n!. \tag{8}$$

Multiply this vector space into the vector space of the preceding section. A general vector in the product space will have co-ordinates A_{nrst} which can be represented as the coefficients in a power series

$$Q = \Sigma_0^\infty A_{nrst}\xi_0^{-n-1}\xi_1^r\xi_2^s\xi_3^t \tag{9}$$

in the four variables ξ_0, ξ_1, ξ_2, ξ_3. The squared length of such a vector is

$$\Sigma_0^\infty n!^{-1} r!\,s!\,t!\,A_{nrst}^2, \tag{10}$$

and two vectors with co-ordinates A_{nrst} and B_{nrst} have a scalar product

$$\Sigma_0^\infty n!^{-1} r!\,s!\,t!\,A_{nrst}B_{nrst}. \tag{11}$$

The series (7) may be extended backwards to include some terms with non-negative powers of ξ_0, so that coefficients k_n occur with negative n-values, leading to coefficients A_{nrst} with negative n-values. Since $n!$ is infinite for n negative, these new coefficients do not contribute to the squared length of a vector or the scalar product of two vectors. Thus the terms with non-negative powers of ξ_0 should be counted as corresponding to the vector zero, and whether they are present in an expansion or not does not matter.

Now apply a Lorentz transformation to the ξ's,

$$\xi_\mu = \alpha_\mu^\nu \xi'_\nu, \tag{12}$$

the α's satisfying certain conditions so that $\xi_0^2 - \xi_1^2 - \xi_2^2 - \xi_3^2$ is invariant. This makes $\xi_1^r \xi_2^s \xi_3^t$ go over into a finite polynomial in ξ'_0, ξ'_1, ξ'_2, ξ'_3, and ξ_0^{-n-1} go over into

$$(\alpha_0^0 \xi'_0 + \alpha_0^1 \xi'_1 + \alpha_0^2 \xi'_2 + \alpha_0^3 \xi'_3)^{-n-1}, \tag{13}$$

which may be expanded in ascending powers of ξ'_1, ξ'_2, ξ'_3 and descending powers of ξ'_0. (The question of the convergence of this expansion is left to the next section, so as not to break the main argument here.) The power series (9) then goes over into a series

$$Q = \Sigma A'_{nrst} \xi_0'^{-n-1} \xi_1'^r \xi_2'^s \xi_3'^t \tag{14}$$

in ascending powers of ξ_1, ξ_2, ξ_3 and descending powers of ξ_0, with coefficients A' which are linear functions of the previous coefficients A. There may be terms in (14) with non-negative powers of ξ_0, but on account of what was said above these can be discarded. If such terms are present in the original series (9), they will not affect any of the coefficients with non-negative n-values in (14).

The Lorentz transformation thus generates a linear transformation of the coefficients A_{nrst} with non-negative n-values. The theorem will now be proved: *The transformation of the coefficients A leaves the squared length* (10) *invariant.*

Consider first the infinitesimal Lorentz transformation

$$\xi_0 = \xi'_0 + \epsilon \xi'_1, \quad \xi_1 = \xi'_1 + \epsilon \xi'_0, \quad \xi_2 = \xi'_2, \quad \xi_3 = \xi'_3. \tag{15}$$

Substituting into (9), one gets

$$Q = \Sigma A_{nrst} [\xi_0'^{-n-1} \xi_1'^r - (n+1)\, \epsilon \xi_0'^{-n-2} \xi_1'^{r+1} + r\epsilon \xi_0'^{-n} \xi_1'^{r-1}]\, \xi_2'^s \xi_3'^t.$$

Hence $$A'_{nrst} = A_{nrst} - n\epsilon A_{n-1,r-1,s,t} + (r+1)\, \epsilon A_{n+1,r+1,s,t},$$

in which A_{nrst} with a negative value for r, s or t is counted as zero. Thus

$$\begin{aligned}&\Sigma n!^{-1} r!\, s!\, t!\, A'^2_{nrst} \\ &\qquad = \Sigma n!^{-1} r!\, s!\, t!\, [A^2_{nrst} - 2n\epsilon A_{nrst} A_{n-1,r-1,s,t} + 2(r+1)\, \epsilon A_{n+1,r+1,s,t} A_{nrst}].\end{aligned}$$

The last two terms in the [ ] here cancel, as may be seen by substituting $n+1$ for n in the former and $r-1$ for r in the latter, and hence the squared length (10) is invariant for the transformation (15). Any Lorentz transformation can be built up from infinitesimal transformations like (15) and three-dimensional rotations like those considered in the preceding section, together with possibly a reflexion, which obviously leaves the squared length (10) invariant, and hence the theorem is proved.

The transformations of the coefficients A thus provide a unitary representation of the Lorentz group. The coefficients themselves form the components of a new kind of tensor quantity in space-time. I propose for it the name *expansor*, because of its connexion with binomial expansions. One may restrict the function (9) to be homogeneous and one then gets a simpler kind of expansor, which may be called a *homogeneous expansor*. The analogy with the three-dimensional case suggests that one should look upon a homogeneous expansor as a symmetrical tensor in space-time with the suffix 0 occurring in its components a negative number of times.

The foregoing theory can be generalized, like the three-dimensional theory, by the introduction of other quadruplets of variables, η_0, η_1, η_2, η_3 and ζ_0, ζ_1, ζ_2, ζ_3 say, which transform together with ξ_0, ξ_1, ξ_2, ξ_3. One can then set up a power series in all the variables, ascending in those variables with suffixes 1, 2 and 3, and descending in those variables with suffix 0. The transformations of the coefficients of these more general power series will provide further unitary representations of the Lorentz group, and the coefficients themselves will form the components of more general expansors.

There is another generalization which may readily be made in the theory, namely, to take the values of the index n in (9) to be not integers, but any set of real numbers n_0, n_0+1, n_0+2, ... extending to infinity. In the formula (10) for the squared length $n!$ is then to be interpreted as $\Gamma(n+1)$. The terms with negative n-values can no longer be discarded. The expression for the squared length is still positive definite if the minimum value of n is greater than -1, which is the case if the function (9) is homogeneous and its degree is negative. The resulting representation is then still unitary. If, however, the function (9) is homogeneous and its degree is positive, there will be a finite number of negative terms in the expression for the squared length. The resulting kind of representation may be called *nearly unitary*.

4. SOME THEOREMS ON CONVERGENCE

If the series (2) is convergent, then (1) is convergent for all values of ξ_1. Similarly, if (4) is convergent, then (3) is convergent for all values of ξ_1, ξ_2, ξ_3. On the other hand, if (8) is convergent, (7) need not be convergent for any value of ξ_0, in which case, of course, it does not define a function of ξ_0. Similarly, if (10) is convergent, (9) need not be convergent for any values for the ξ's. Thus, corresponding to a general expansor A_{nrst}, there need not exist any function Q of the ξ's. However, it will now be proved that *if the series* (9) *is homogeneous and* (10) *is convergent, then* (9) *is absolutely convergent for all values of the ξ's satisfying*

$$\xi_0^2-\xi_1^2-\xi_2^2-\xi_3^2>0. \tag{16}$$

If the series (9) is homogeneous of degree $u-1$, it may be written

$$\Sigma_n \xi_0^{-n-1} \Sigma_S A_{nrst} \xi_1^r \xi_2^s \xi_3^t, \tag{17}$$

where Σ_S means a sum over all values of r, s and t satisfying

$$r+s+t = n+u. \tag{18}$$

With this notation the series (10) may be written

$$\Sigma_n n!^{-1} \Sigma_S r!\,s!\,t!\,A_{nrst}^2. \tag{19}$$

Now apply Cauchy's inequality

$$(x_1y_1+x_2y_2+x_3y_3+\dots)^2 \leqslant (x_1^2+x_2^2+x_3^2+\dots)(y_1^2+y_2^2+y_3^2+\dots) \tag{20}$$

in the following way. Take $(r+s+t)!/r!s!t!$ of the x's in (20) to be each equal to $\xi_1^r\xi_2^s\xi_3^t$ and the corresponding y's to be each equal to $A_{nrst}r!s!t!/(r+s+t)!$, and do this for all r, s, t subject to (18) for a fixed value of n. Then (20) becomes

$$(\Sigma_S A_{nrst}\xi_1^r\xi_2^s\xi_3^t)^2 \leqslant (\xi_1^2+\xi_2^2+\xi_3^2)^{n+u}(n+u)!^{-1}\Sigma_S r!s!t!A_{nrst}^2. \tag{21}$$

If the sum with respect to n in (19) converges, there must be some number κ such that

$$n!^{-1}\Sigma_S r!s!t!A_{nrst}^2 < \kappa$$

for all n. Then from (21)

$$|\xi_0^{-n-1}\Sigma_S A_{nrst}\xi_1^r\xi_2^s\xi_3^t| < \kappa^{\frac{1}{2}}|\xi_0|^{u-1}\left\{\frac{n!}{(n+u)!}\right\}^{\frac{1}{2}}\left\{\frac{\xi_1^2+\xi_2^2+\xi_3^2}{\xi_0^2}\right\}^{\frac{1}{2}(n+u)},$$

which shows that (17) is convergent when (16) is satisfied. One may take ξ_0, ξ_1, ξ_2, ξ_3 and the A's to be all positive without disturbing the argument, so (17), considered as a quadruple series in n, r, s, t, is absolutely convergent.

Thus for any homogeneous expansor of finite length, there exists a function Q given by (9) defined within the light-cone (16). In transforming this function with a Lorentz transformation (12), it is legitimate to use the expansion of (13) in ascending powers of ξ_1', ξ_2', ξ_3' and descending powers of ξ_0', for the following reason. Suppose ξ_μ is within the light-cone and take for definiteness $\xi_0>0$. Then

$$\alpha_0^0\xi_0'+\alpha_0^1\xi_1'+\alpha_0^2\xi_2'+\alpha_0^3\xi_3'>0. \tag{22}$$

One can change the sign of any of the co-ordinates ξ_1', ξ_2', ξ_3', leaving ξ_0' unchanged, and ξ_μ will still lie within the light-cone with $\xi_0>0$, so (22) must still be satisfied. Hence

$$\alpha_0^0\xi_0'-|\alpha_0^1\xi_1'|-|\alpha_0^2\xi_2'|-|\alpha_0^3\xi_3'|>0,$$

which shows that the expansion of (13) in the required manner is absolutely convergent. The use of this expansion in the preceding section is thus justified for the case when (9) is homogeneous, the new coefficients A being determined by the transformed function Q within the light-cone. The justification for (9) not homogeneous then follows since terms of different degree do not interfere. It should be noted that all the foregoing arguments are valid also when the values of n are not integers.

There are some expansors which are invariant under all Lorentz transformations, namely those whose components are the coefficients of $(\xi_0^2-\xi_1^2-\xi_2^2-\xi_3^2)^{-l}$ expanded in ascending powers of ξ_1, ξ_2, ξ_3 and descending powers of ξ_0. For such an expansion to be possible, l must not be a negative integer or zero, but it can be any other real number. Again, there are expansors which transform like ordinary tensors under Lorentz transformations, namely those whose components are the coefficients in the expansion of

$$f(\xi_0,\xi_1,\xi_2,\xi_3)(\xi_0^2-\xi_1^2-\xi_2^2-\xi_3^2)^{-l}, \tag{23}$$

where f is any homogeneous integral polynomial in ξ_0, ξ_1, ξ_2, ξ_3, and l is restricted as before. (23) transforms like the polynomial f, and thus like a tensor of order equal to the degree of f. One would expect all these expansors to be of infinite length, as otherwise one could set up a positive definite form for the squared length of a tensor in space-time. A formal proof that they are is as follows.

Suppose the f in (23) is of degree u and express it as

$$f = g_u + \xi_0 g_{u-1} + (\xi_0^2 - \boldsymbol{\xi}^2)\, g_{u-2} + \xi_0(\xi_0^2 - \boldsymbol{\xi}^2)\, g_{u-3} + (\xi_0^2 - \boldsymbol{\xi}^2)^2 g_{u-4} + \ldots,$$

where $\boldsymbol{\xi}^2 = \xi_1^2 + \xi_2^2 + \xi_3^2$ and each of the g's is a polynomial in ξ_1, ξ_2, ξ_3 only, of degree indicated by the suffix. The successive terms here contribute to (23) the amounts

$$\begin{aligned}
g_u(\xi_0^2 - \boldsymbol{\xi}^2)^{-l} &= \Sigma_{m=0}^{\infty} \frac{l(l+1)(l+2)\ldots(l+m-1)}{m!} \frac{g_u \boldsymbol{\xi}^{2m}}{\xi_0^{2(m+l)}}, \\
g_{u-1}\xi_0(\xi_0^2 - \boldsymbol{\xi}^2)^{-l} &= \Sigma_{m=0}^{\infty} \frac{l(l+1)(l+2)\ldots(l+m-1)}{m!} \frac{g_{u-1} \boldsymbol{\xi}^{2m}}{\xi_0^{2(m+l)-1}}, \\
g_{u-2}(\xi_0^2 - \boldsymbol{\xi}^2)^{-l+1} &= \Sigma_{m=-1}^{\infty} \frac{(l-1)\,l(l+1)\ldots(l+m-1)}{(m+1)!} \frac{g_{u-2} \boldsymbol{\xi}^{2(m+1)}}{\xi_0^{2(m+l)}}, \\
g_{u-3}\xi_0(\xi_0^2 - \boldsymbol{\xi}^2)^{-l+1} &= \Sigma_{m=-1}^{\infty} \frac{(l-1)\,l(l+1)\ldots(l+m-1)}{(m+1)!} \frac{g_{u-3} \boldsymbol{\xi}^{2(m+1)}}{\xi_0^{2(m+l)-1}},
\end{aligned}$$

and so on. These expansions show that, for large values of m, the terms arising from g_{u-2}, g_{u-4}, ... are of smaller order than the corresponding terms arising from g_u, and the terms arising from g_{u-3}, g_{u-5}, ... are of smaller order than the corresponding terms arising from g_{u-1}, so that in testing for the convergence of the series which gives the squared length, only the g_u and g_{u-1} terms need be taken into account. (It will be found that the convergence conditions are not sufficiently critical to be affected by this neglect.) Now express g_u and g_{u-1} as

$$g_u = S_u + \boldsymbol{\xi}^2 S_{u-2} + \boldsymbol{\xi}^4 S_{u-4} + \ldots, \quad g_{u-1} = S_{u-1} + \boldsymbol{\xi}^2 S_{u-3} + \boldsymbol{\xi}^4 S_{u-5} + \ldots,$$

where the S's are solid harmonic functions of ξ_1, ξ_2, ξ_3, of degrees indicated by the suffixes. Each of them gives a contribution to (23) of the form

$$S_{u-2r}\boldsymbol{\xi}^{2r}(\xi_0^2 - \boldsymbol{\xi}^2)^{-l} = S_{u-2r}\, \Sigma_{m=0}^{\infty} \frac{(m+l-1)!}{m!(l-1)!} \frac{\boldsymbol{\xi}^{2(m+r)}}{\xi_0^{2(m+l)}} \tag{24}$$

or

$$S_{u-2r-1}\xi_0\boldsymbol{\xi}^{2r}(\xi_0^2 - \boldsymbol{\xi}^2)^{-l} = S_{u-2r-1}\, \Sigma_{m=0}^{\infty} \frac{(m+l-1)!}{m!(l-1)!} \frac{\boldsymbol{\xi}^{2(m+r)}}{\xi_0^{2(m+l)-1}}. \tag{25}$$

Using the result (41) of the appendix, one finds for the squared lengths of the expansors whose components are the coefficients of (24) and (25), series of the form

$$c\,\Sigma_m \frac{(m+l-1)!^2}{m!^2(l-1)!^2} \cdot \frac{4^{m+r}(m+r)!\,(m+u-r+\tfrac{1}{2})!}{(2m+2l-1)!}$$

and

$$c\,\Sigma_m \frac{(m+l-1)!^2}{m!^2(l-1)!^2} \frac{4^{m+r}(m+r)!\,(m+u-r-\tfrac{1}{2})!}{(2m+2l-2)!},$$

respectively. In both these series the ratio of the $(m+1)$th term to the mth is $1+u/m$ for large m, and since u is necessarily positive or zero, the series diverge and the expansors are of infinite length. From the result (40) of the appendix, expansors associated with solid harmonics of different degrees are orthogonal to one another, and hence the total expansor with the coefficients of (23) as components must also be of infinite length.

5. A TRANSFORMATION OF VARIABLES

Up to the present, expansors have always been considered in connexion with certain ξ-functions, the components of the expansor being the coefficients of the ξ-function. In the case of integral n-values, one can make a transformation of variables of the kind which is familiar in quantum mechanics, and get the expansors connected with some other functions, which serves to clarify some of their properties.

Introduce the four operators x_μ, defined by

$$2^{\frac{1}{2}}x_0 = \xi_0 - \partial/\partial\xi_0, \quad 2^{\frac{1}{2}}x_r = \xi_r + \partial/\partial\xi_r \quad (r = 1, 2, 3). \tag{26}$$

Under Lorentz transformations they transform like the components of a 4-vector. The operators $\partial/\partial x_\mu$ may be taken to be

$$2^{\frac{1}{2}}\,\partial/\partial x_0 = \xi_0 + \partial/\partial\xi_0, \quad 2^{\frac{1}{2}}\,\partial/\partial x_r = -\xi_r + \partial/\partial\xi_r, \tag{27}$$

as this gives the correct commutation relations between all the operators. One can now represent ξ-functions with integral n-values by functions of the x's according to the following scheme.

Take as starting point the ξ-function ξ_0^{-1}. It vanishes when operated on by $\partial/\partial\xi_1$, $\partial/\partial\xi_2$, or $\partial/\partial\xi_3$, and also effectively vanishes when multiplied by ξ_0, since the result of the multiplication can be discarded. It must therefore be represented by a function of the x's which vanishes when operated on by $x_r + \partial/\partial x_r$ with $r = 1, 2$ or 3, or by $x_0 + \partial/\partial x_0$, and thus by a multiple of $e^{-\frac{1}{2}(x_0^2+\mathbf{x}^2)}$, where $\mathbf{x}^2 = x_1^2 + x_2^2 + x_3^2$. Take

$$\xi_0^{-1} \equiv \pi^{-1} e^{-\frac{1}{2}(x_0^2+\mathbf{x}^2)}, \tag{28}$$

where the sign $\equiv$ means 'represented by'. Multiply the left-hand side of (28) by the operator $(-\partial/\partial\xi_0)^n\,\xi_1^r\xi_2^s\xi_3^t$ and the right-hand side by the operator equal to it according to (26) and (27). The result is, after dividing through by $n!$,

$$\begin{aligned}\xi_0^{-n-1}\xi_1^r\xi_2^s\xi_3^t &\equiv \pi^{-1}n!^{-1}\,2^{-\frac{1}{2}(n+r+s+t)}\left(x_0 - \frac{\partial}{\partial x_0}\right)^n \left(x_1 - \frac{\partial}{\partial x_1}\right)^r \left(x_2 - \frac{\partial}{\partial x_2}\right)^s \left(x_3 - \frac{\partial}{\partial x_3}\right)^t e^{-\frac{1}{2}(x_0^2+\mathbf{x}^2)} \\ &= F_{nrst}(x_0, x_1, x_2, x_3),\end{aligned} \tag{29}$$

say. This gives the function F of the x's which represents $\xi_0^{-n-1}\xi_1^r\xi_2^s\xi_3^t$. A general ξ-function is now represented thus,

$$\Sigma A_{nrst}\,\xi_0^{-n-1}\xi_1^r\xi_2^s\xi_3^t \equiv \Sigma A_{nrst}F_{nrst}(x_0, x_1, x_2, x_3). \tag{30}$$

In this way a general expansor A_{nrst} gets connected with a function of the x's.

The chief interest of this connexion is that the law for the scalar product of two expansors becomes very simple when expressed in terms of the functions of the x's. *The scalar product is the integral of the product of the two functions of the x's over all* x_0, x_1, x_2 *and* x_3. To prove this, first evaluate the integral

$$\int_{-\infty}^{\infty}\left(z-\frac{d}{dz}\right)^m e^{-\frac{1}{2}z^2}.\left(z-\frac{d}{dz}\right)^{m'} e^{-\frac{1}{2}z^2}\,dz, \tag{31}$$

where the dot has the meaning that operators to the left of it do not operate on functions to the right of it. For $m>0$, (31) goes over by partial integration into

$$\begin{aligned}&\int_{-\infty}^{\infty}\left(z-\frac{d}{dz}\right)^{m-1} e^{-\frac{1}{2}z^2}.\left(z+\frac{d}{dz}\right)\left(z-\frac{d}{dz}\right)^{m'} e^{-\frac{1}{2}z^2}\,dz\\ &=\int_{-\infty}^{\infty}\left(z-\frac{d}{dz}\right)^{m-1} e^{-\frac{1}{2}z^2}.\left\{\left(z-\frac{d}{dz}\right)^{m'}\left(z+\frac{d}{dz}\right)+2m'\left(z-\frac{d}{dz}\right)^{m'-1}\right\} e^{-\frac{1}{2}z^2}\,dz\\ &=2m'\int_{-\infty}^{\infty}\left(z-\frac{d}{dz}\right)^{m-1} e^{-\frac{1}{2}z^2}.\left(z-\frac{d}{dz}\right)^{m'-1} e^{-\frac{1}{2}z^2}\,dz.\end{aligned}$$

Applying this procedure m times, one gets zero if $m'<m$ and $2^m m!\,\pi^{\frac{1}{2}}$ if $m'=m$. Since there is symmetry between m and m', the result is

$$\int_{-\infty}^{\infty}\left(z-\frac{d}{dz}\right)^m e^{-\frac{1}{2}z^2}.\left(z-\frac{d}{dz}\right)^{m'} e^{-\frac{1}{2}z^2}\,dz=2^m m!\,\pi^{\frac{1}{2}}\delta_{mm'}.$$

Now substitute for z each of the variables x_0, x_1, x_2, x_3 in turn, with m equal to n, r, s, t and m' equal to n', r', s', t' respectively, and multiply the four equations so obtained. The result, after dividing through by $\pi^2 n!^2\, 2^{n+r+s+t}$, is

$$\iiiint F_{nrst}F_{n'r's't'}\,dx_0\,dx_1\,dx_2\,dx_3=n!^{-1}r!\,s!\,t!\,\delta_{nn'}\delta_{rr'}\delta_{ss'}\delta_{tt'}.$$

This proves the theorem for the case when the expansors each have only one non-vanishing component, which is sufficient to prove it generally.

One can obviously get a unitary representation of the Lorentz group by considering the transformations of a set of vectors in a space of an infinite number of dimensions, where each vector corresponds to a function of four Lorentz variables x_0, x_1, x_2, x_3, and where the squared length of a vector is defined as the integral of the square of the function over all x_0, x_1, x_2, x_3. The above work shows how this obvious unitary representation is connected with the expansor theory.

6. APPLICATIONS TO QUANTUM MECHANICS

The four x's of the preceding section may be looked upon as the co-ordinates of a four-dimensional harmonic oscillator, the four operators $i\partial/\partial x_\mu$ being the conjugate momenta p^μ, and the energy of the oscillator may be taken to be the Lorentz invariant

$$\tfrac{1}{2}[x_1^2+x_2^2+x_3^2-x_0^2+(p^1)^2+(p^2)^2+(p^3)^2-(p^0)^2]. \tag{32}$$

The 1, 2, 3 components of the oscillator thus have positive energies and the 0 component negative energy.

The state of the oscillator for which the 0, 1, 2, 3 components are in the nth, rth, sth, tth quantum states respectively is then represented by the function F_{nrst} defined by (29), with a suitable normalizing factor. This representative may be transformed to the ξ-representation and becomes $\xi_0^{-n-1}\xi_1^r\xi_2^s\xi_3^t$. Thus a state of the oscillator for which each of its components is in a quantum state corresponds to an expansor with one non-vanishing component. A general state of the oscillator therefore corresponds to a general expansor with integral n-values. A stationary state of the oscillator corresponds to a homogeneous expansor, the degree of the expansor giving the energy of the state with neglect of zero-point energy.

Four-dimensional harmonic oscillators of the above type occur in the theory of the electromagnetic field. Each Fourier component of the field, specified by a particular frequency and a particular direction of motion of the waves, provides one such oscillator, its four components coming from the four electromagnetic potentials. Thus a state of the electromagnetic field in quantum mechanics is described by a number of expansors, one for each Fourier component. By using the electromagnetic equation which gives the value of the divergence of the potentials, one can eliminate in a non-relativistic way the 0 component and one other component of each of the four-dimensional oscillators, so that only two-dimensional oscillators are left. This circumstance has made it possible for people to develop quantum electrodynamics without using expansors.

Another possible application of expansors is to the spins of particles. The wave function describing a particle may be a function of the four co-ordinates x_μ of the particle in space-time and also of the four variables ξ_μ whose coefficients are the components of an expansor. As a simple example of relativistic wave equations for such a particle in the absence of external forces, one may consider

$$\left(\hbar^2\frac{\partial^2}{\partial x_\mu \partial x^\mu}+m^2\right)\psi = 0, \quad \xi_\mu\frac{\partial}{\partial \xi_\mu}\psi = -\psi, \quad \xi_\mu\frac{\partial}{\partial x_\mu}\psi = 0. \tag{33}$$

The first of these is the usual equation for the motion of the particle as a whole. The second shows that at each point in space-time the wave function ψ is homogeneous in the ξ's of degree -1. The third shows that the state for which the momentum-energy four-vector of the particle has the value p_μ is represented by the wave function

$$\psi = \frac{m}{\xi_0 p_0 - (\xi\mathbf{p})} e^{-i[p_0x_0-(\mathbf{px})]/\hbar}$$

in three-dimensional vector notation. This may be expanded as

$$\psi = me^{-i[p_0x_0-(\mathbf{px})]/\hbar}\sum_{n=0}^{\infty}\frac{(\xi\mathbf{p})^n}{\xi_0^{n+1}p_0^{n+1}}. \tag{34}$$

For the state for which the particle is at rest, $p_1 = p_2 = p_3 = 0$, $p_0 = m$, and ψ reduces to

$$\psi = e^{-imx_0/\hbar}\xi_0^{-1}.$$

This ψ is spherically symmetrical, showing that when the particle is at rest it has no spin. But when the particle is moving, it is represented by the general ψ (34) and has a finite probability of a non-zero spin. In fact, taking for simplicity $p_2 = p_3 = 0$, the particle has a probability $(mp_1^n/p_0^{n+1})^2$ of being in a state of spin corresponding to the transformations of ξ_1^n under three-dimensional rotations.

This example shows there is a possibility of a particle having no spin when at rest but acquiring a spin when moving, a state of affairs which was not allowed by previous theory. It is desirable that the new spin possibilities opened up by the present theory should be investigated to see whether they could in some cases give an improved description of Nature. The present theory of expansors applies, of course, only to integral spins, but probably it will be possible to set up a corresponding theory of two-valued representations of the Lorentz group, which will apply to half odd integral spins.

APPENDIX

The rules (5), (11) for forming scalar products are not always convenient for direct use. There are various ways of transforming them and making them more suitable for practical application. One such way has been given (Dirac 1942, equation 3·22) for the case of a single ξ with ascending power series. Another way, applicable to the case of homogeneous functions of ξ_1, ξ_2, ξ_3, is provided by the following.

By partial integration with respect to ξ', one gets, for $m > 0$,

$$\iint_{-\infty}^{\infty} \xi^m \xi'^n e^{i\xi\xi'}\, d\xi\, d\xi' = \int_{-\infty}^{\infty} \xi^{m-1}\, d\xi \left\{ \Big[-i\xi'^n e^{i\xi\xi'} \Big]_{\xi'=-\infty}^{\xi'=\infty} + in \int_{-\infty}^{\infty} \xi'^{n-1} e^{i\xi\xi'}\, d\xi' \right\}.$$

If the integrals are made precise in the sense of Cesaro, which means neglecting oscillating terms like $\xi'^n e^{i\xi\xi'}$ for ξ' infinite, this gives

$$\iint_{-\infty}^{\infty} \xi^m \xi'^n e^{i\xi\xi'}\, d\xi\, d\xi' = in \iint_{-\infty}^{\infty} \xi^{m-1} \xi'^{n-1} e^{i\xi\xi'}\, d\xi\, d\xi'.$$

Taking $m \geqslant n$ and applying the partial integration process n times, one gets

$$\begin{aligned} \iint_{-\infty}^{\infty} \xi^m \xi'^n e^{i\xi\xi'}\, d\xi\, d\xi' &= i^n n! \iint_{-\infty}^{\infty} \xi^{m-n} e^{i\xi\xi'}\, d\xi\, d\xi' \\ &= 2\pi i^n n! \int_{-\infty}^{\infty} \xi^{m-n}\, \delta(\xi)\, d\xi \\ &= 2\pi i^n n!\, \delta_{mn}. \end{aligned} \tag{35}$$

It follows that if A and B are homogeneous functions of ξ_1, ξ_2, ξ_3 of degree u, their scalar product according to (5) is

$$(AB) = (2\pi)^{-3} i^{-u} \iint_{\cdots\cdots} A(\xi_1 \xi_2 \xi_3)\, B(\xi_1' \xi_2' \xi_3')\, e^{i(\xi_1\xi_1' + \xi_2\xi_2' + \xi_3\xi_3')}\, d\xi_1\, d\xi_1'\, d\xi_2\, d\xi_2'\, d\xi_3\, d\xi_3'. \tag{36}$$

As an application of this rule, take

$$A = (\xi_1^2+\xi_2^2+\xi_3^2)^r S_{u-2r}, \quad B = (\xi_1^2+\xi_2^2+\xi_3^2)^s S_{u-2s},$$

where the S's are solid harmonic functions. Then, using three-dimensional vector notation, (36) gives

$$(AB) = (2\pi)^{-3} i^{-u} \iint_{\cdots\cdots} \xi^{2r}\xi'^{2s} S_{u-2r}(\xi)\, S_{u-2s}(\xi')\, e^{i(\xi\xi')}\, d\xi_1 \ldots d\xi_3'. \tag{37}$$

From Green's theorem

$$\iiint\left[e^{i(\xi\xi')}\left(\frac{\partial^2}{\partial\xi_1^2}+\frac{\partial^2}{\partial\xi_2^2}+\frac{\partial^2}{\partial\xi_3^2}\right)\left\{\xi^{2r}S_{u-2r}(\xi)\right\}-\xi^{2r}S_{u-2r}(\xi)\left(\frac{\partial^2}{\partial\xi_1^2}+\frac{\partial^2}{\partial\xi_2^2}+\frac{\partial^2}{\partial\xi_3^2}\right)e^{i(\xi\xi')}\right]d\xi_1 d\xi_2 d\xi_3$$

equals a surface integral of an oscillating kind which is to be counted as vanishing at infinity. This result reduces to

$$4r(u-r+\tfrac{1}{2})\iiint_{-\infty}^{\infty}\xi^{2(r-1)}S_{u-2r}(\xi)\, e^{i(\xi\xi')}\, d\xi_1 d\xi_2 d\xi_3+\xi'^2\iiint_{-\infty}^{\infty}\xi^{2r}S_{u-2r}(\xi)\, e^{i(\xi\xi')}\, d\xi_1 d\xi_2 d\xi_3 = 0.$$

so (37) becomes, for $r, s > 0$,

$$(AB) = (2\pi)^{-3} i^{-u+2}\, 4r(u-r+\tfrac{1}{2}) \iint_{\cdots\cdots} \xi^{2(r-1)}\xi'^{2(s-1)} S_{u-2r}(\xi)\, S_{u-2s}(\xi')\, e^{i(\xi\xi')}\, d\xi_1 \ldots d\xi_3'. \tag{38}$$

Now suppose $s \geqslant r$ and apply the procedure by which (37) was changed to (38) r times. The result is

$$(AB) = (2\pi)^{-3} i^{-u+2r}\, 4^r r!\,(u-r+\tfrac{1}{2})!\,(u-2r+\tfrac{1}{2})!^{-1} \times \iint_{\cdots\cdots} \xi'^{2(s-r)} S_{u-2r}(\xi) S_{u-2s}(\xi')\, e^{i(\xi\xi')}\, d\xi_1 \ldots d\xi_3'. \tag{39}$$

where $n!$ means $\Gamma(n+1)$ for n not an integer. If $s > r$, the procedure can be applied once more, and then shows that

$$(AB) = 0 \quad \text{for} \quad r \neq s. \tag{40}$$

If $s = r$, (39) shows that $$(AB) = c4^r r!\,(u-r+\tfrac{1}{2})!, \tag{41}$$

where c depends only on $u-2r$ and on the two S functions.

REFERENCE

Dirac, P. A. M. 1942 *Proc. Roy. Soc.* A, **180**, 1–40.

An investigation of the performance of the Rayleigh disk

By R. A. Scott, *Research Department, Metropolitan Vickers Electrical Co. Ltd.*

(*Communicated by Professor D. R. Hartree, F.R.S.—Received* 2 *May* 1944)

The paper describes an investigation of the nature of the law which governs the torque on a Rayleigh disk suspended in a sound field. The torque is found to follow an expression of the form obtained theoretically by König, provided that factors such as the finite thickness of the disk and the lack of infinite inertia of the disk are taken into account. The measurements were made at frequencies spread over the range 250–4000 cyc./sec. The correction to the torque in König's formula made necessary by the thickness of the disk is shown to be large and to have a sign opposite to that determined theoretically by König for ellipsoidal disks. The influence upon the torque of the proximity of the wall of a tube has been investigated. The nature of the precautions which should be observed in choosing the form of a Rayleigh disk are discussed with reference to reducing to a minimum the effects on the torque of the thickness and mobility of the disk and of diffraction of sound by the disk.

An investigation has been made of the numerical factor which relates the magnitudes of the torque to the other variables of König's formula. This part of the investigation reduces essentially to the measurement of the oscillatory velocity of the field in which the disk is situated. The smoke-particle method of Andrade and of Carrière was adapted for the above purpose, and it was found possible to make determinations of oscillatory velocity within about 1 % for frequencies up to 4000 cyc./sec.

Measurements have been made of the torque on a Rayleigh disk of known dimensions in terms of the oscillatory velocity for frequencies in the range 250–4000 cyc./sec. Discrepancies of 3·5 % in velocity (7 % in torque) are shown to follow from the use of König's formula, and these results confirm, and extend into the audio-frequency range, the investigations of Merrington & Oatley in the band 9–22 cyc./sec.

The experimental work confirms that the stability of behaviour of the Rayleigh disk justifies its continued use as a reference standard of acoustical intensity but shows that in setting up the reference standard, corrections amounting in extreme cases to 0·3 db. must be applied to the intensity as determined from König's formula.

1. Introduction

The Rayleigh disk has long received attention as a means of providing a standard of sound intensity. Its simplicity of construction, and the ease with which it can be used under suitable experimental conditions, alike make it attractive and have led to its common use as a standard against which to calibrate microphones. Rayleigh (1880, 1882) showed that a torque is exerted on a thin circular disk when immersed in a flowing medium. The torque tends to turn the disk at right angles to the direction of flow. König (1891) determined mathematically the torque on such a disk on the assumption that the flow is streamline and steady. He found that an infinitely thin disk of radius a, when set so that its normal makes an angle θ with the direction of flow, should experience a torque given by the expression

$$L = \tfrac{4}{3}\rho a^3 V^2 \sin 2\theta, \tag{1}$$

where ρ is the density of the medium (air) and V is the velocity of the flow. König showed that the effect of viscosity of the medium is, to a first approximation, to produce a thrust and not a torque. The expression (1) shows that the torque remains unchanged in sign when the velocity of the medium is reversed, and it is for this reason that the Rayleigh disk finds application in acoustics.

When a disk is placed in a field of oscillatory velocity, L and V must be taken to stand for the mean value of the torque and the root-mean-square velocity respectively, and König's equation in this form serves as a basis of acoustical measurement. Various investigators (e.g. Zernov 1908; Barnes & West 1927) have examined the applicability of König's formula to acoustic fields, but almost all of the work has been carried out at the very lowest end of the audio-frequency range. Until recently no serious fault has been found with the formula, apart from small corrections which have been necessary for the effect of the finite thickness and the lack of complete immobility of the disk. Merrington & Oatley (1939) observed, however, whilst working at very low frequency (9–22 cyc./sec.), a discrepancy of 10 % between the torque on the disk and that deduced from König's formula. This discrepancy was attributed to such steady circulations of air about the disk as are under given conditions to be found around obstacles in a sound field (Andrade 1931). Merrington & Oatley demonstrated that circulations of the appropriate type are to be found about the Rayleigh disk under the conditions of their experiments.

The discovery that the torque on a Rayleigh disk as calculated from König's formula can be in error at low frequency by as much as 10 % throws doubt on calibrations made at all frequencies. It is therefore important to investigate the accuracy of König's formula over the entire range of audio-frequency. In the present paper the behaviour of the Rayleigh disk over the range of frequency 250–4250 cyc./sec. is investigated. The general form of the expression for the torque as found experimentally is described in the first part of the work, whilst in the second part an account is given of an experimental comparison between oscillatory velocity as determined on the one hand by application of König's formula to the Rayleigh disk, and on the other hand from the excursions of smoke particles entrained in the sound field.

2. DETERMINATION OF THE FORM OF THE LAW GOVERNING THE TORQUE ON THE RAYLEIGH DISK

Before König's formula is applied to the case of an actual Rayleigh disk, the influence on the torque of the following factors, which were not adequately dealt with in König's analysis, must be investigated:

(*a*) The finite thickness of the disk.

(*b*) The lack of infinite inertia of the disk.

(*c*) The diffraction of the sound field by the disk.

(*d*) The restriction upon the streaming near the disk caused by the proximity of the wall of any enclosing tube.

Each of these factors has in the past received some attention, but only (*b*) has been investigated satisfactorily. The corrections are by no means negligible, and were therefore investigated in detail.

(a) *Finite thickness of disk*

No theoretical expression has yet been obtained for the torque on a short cylinder immersed in a sound field. König's paper includes a derivation of the following expression for the torque on an ellipsoid of circular principal section of radius a:

$$L = \tfrac{4}{3}\rho a^3 V^2 \sin 2\theta(1 - 0{\cdot}30x), \tag{2}$$

provided that x, the ratio of thickness to diameter, is much less than unity. This expression implies that an increase in thickness of the disk leads to a decrease in torque on the disk for disks of ellipsoidal shape. Zernov (1908) has found experimentally that the torque on disks of short cylindrical form is given by

$$L = L_0(1 + 2{\cdot}78x - 9{\cdot}05x^2), \tag{3}$$

where $L_0 = \frac{4}{3}\rho a^3 V^2 \sin 2\theta$ and x is the ratio of thickness to diameter of the disk. The measurements were made on a small number of disks at a frequency of about 90 cyc./sec. This expression indicates that an increase in torque accompanies an increase in thickness of the disk, provided that x is not very large and this is at variance with König's theoretical result for ellipsoids. It is clear that the correction for finite thickness of the disk requires further investigation.

(b) *Lack of infinite inertia of the disk, and*
(c) *Diffraction of sound by the disk*

A. B. Wood showed in 1935 that a Rayleigh disk immersed in an accelerating medium follows to some extent the motion of the medium. This motion of the disk leads to a reduction in relative velocity between the disk and the medium and therefore to a decreased torque on the disk. Wood has shown that because of movement of the disk, the torque on a light disk used under water may be reduced as much as 2500-fold from the König value.

L. V. King (1935) has produced an excellent mathematical analysis of the behaviour of the Rayleigh disk and has paid particular attention to the effects of finite inertia of the disk and of diffraction of the sound wave by the disk. For a disk of radius a placed at an antinode in a system of plane stationary waves, King finds

$$L = L_0 \frac{m_1[1 + \frac{2}{5}(ka)^2 \cos^2\theta]}{m_1 + m_0[1 + \frac{1}{5}(ka)^2]}, \tag{4}$$

provided that the disk has a diameter small compared with the wave-length of the sound. In this formula m_1 is the mass of the disk less the mass of the medium displaced by the disk, m_0 is the hydrodynamic mass of the disk ($\frac{8}{3}\rho a^3$), and k is 2π times the reciprocal of the wave-length. At low frequency, where ka is much less than unity, the correction factor reduces, for $\theta = 45°$, to $m_1/(m_1 + m_0)$, and this is in agreement with Wood's expression for the effect and also with the results of his experiments in water. There can be little doubt that this correction should be applied to all experimental values of torque on the disk.

For frequencies at which ka approaches unity the correction factor becomes, for $\theta = 45°$, approximately $(1+\frac{1}{5}(ka)^2 - m_0/m_1)$. This diffraction correction amounts in air to about 1 % of the torque for a disk of $\frac{1}{4}$ in. diameter at 3800 cyc./sec., i.e. for $ka = 0{\cdot}22$. The correction is therefore important only at the higher end of the audio-frequency range.

In most investigations, including our own, in which the Rayleigh disk is used, care is taken to work only with disks of diameter small compared with the wavelength. The smallness of the correction would seem to justify, at least until further evidence is available, the application of King's theoretical correction, since this rests on as firm a basis as does the general form of König's expression for the primary torque. Barnes & West, studying the behaviour of Rayleigh disks at high frequency, find that discrepancies of more than 2 % do not appear between the actual torque and that given by a formula of the same type as König's, provided that the ratio of disk diameter to wave-length is less than 0·2.

(*d*) *Restriction upon the streaming near the disk, caused by the proximity of the walls of the enclosing tube*

Tomotika (1937) has recently calculated the effect of enclosing walls upon the torque in the case of the two-dimensional analogue of the Rayleigh disk and finds the torque to be increased in the factor

$$(1+0{\cdot}411z^2+0{\cdot}247z^4+0{\cdot}120z^6\ldots), \tag{5}$$

where z = (width of disk)/(width of channel), and where $\theta = 45°$. His calculation assumes the medium to be incompressible and therefore ignores wave effects. His paper includes results of measurements on disks made by Sato in support of the calculations. The brevity of the account makes appraisement of the observations difficult. These experimental results are in qualitative agreement with the two-dimensional theory, but show a substantial dependence on frequency.

Zernov (*loc. cit.*) has found experimentally that in the presence of an enclosing tube the torque is increased in the ratio

$$(1+1{\cdot}2z^2-5{\cdot}8z^4):1, \tag{6}$$

where z = (diameter of disk)/(diameter of tube). However, only a small range of z was covered in the work, and the maximum value of the correction was 6·1 %, corresponding to $z = 0{\cdot}3$. Barnes & West (1927), working with disks of various sizes, have observed no marked effect on the torque due to the presence of the walls. It is evident that further investigation of this correction is desirable.

It is appropriate here to draw attention to a factor which has been reported to cause, on occasion, erratic behaviour of the disk (Barnes & West, 1927). When the frequency of the sound is close to a resonant frequency of flexural vibration of the disk, the vibrations of the disk modify the sound field and lead to anomalous behaviour. The anomalies are limited, however, to a small range of frequency close to the resonant frequency of the disk.

In order to determine with more certainty the influence on the torque of the factors enumerated above, especially in the range of frequency 250–4000 cyc./sec., measurements were made by the author to determine how the torque on the disk varies as changes are made in the dimensions and disposition of the disk and in the intensity and frequency of the sound.

3. APPARATUS

(a) *Resonant tube and the Rayleigh disk*

Investigation at audio-frequencies of the torque on Rayleigh disks must be made in a draught-free space and in an acoustic field of known direction and of good uniformity. The most convenient way of providing these conditions is to immerse the disk in the sound field in a resonant tube, such as can be made by closing one end of a tube with a solid reflector and placing a source of sound at the other end. Rayleigh has shown that the wave front inside a long tube tends to become plane and normal to the axis provided that the ratio of the diameter of the tube to the wave-length is less than 0·59. Experimental explorations of the field in a closed-end resonant tube have been made by the author for several tubes with the aid of a microphone which receives the sound via a narrow probe-tube (0·159 cm. bore, 0·079 cm. wall) and it has been found that fields which are to a close approximation plane-stationary can be obtained up to frequencies indicated by the Rayleigh criterion. Marked deviations from the plane-stationary field no not occur until roughly twice this frequency is reached, provided that the attenuation along the tube is not too great.

Subject to the limitations discussed above, the acoustic field in a closed tube consists of plane, stationary waves in which velocity antinodes are spaced at odd multiples of $\frac{1}{4}\lambda$ from the reflector end of the tube. When the tube is in effect closed at both ends and in resonance the total length of the tube is some multiple of the half wave-length (apart from end corrections which are usually small). At all frequencies for which the tube is an odd number of half wave-lengths long, the reflector is at a pressure maximum and the midsection of the length of the tube at a velocity maximum. In the region of the velocity maximum, the velocity is axial and substantially uniform. The midsection of the tube is therefore a suitable place at which to use a Rayleigh disk. The reflector at the end of the tube consists of the flat diaphragm of a microphone. Neglecting dissipation in the tube,* it can be shown that the r.m.s. pressure p at a node in the tube is related to the r.m.s. velocity V at an antinode by the expression

$$p = \rho c V, \tag{7}$$

where ρ is the density of the air and c is the velocity of sound. The e.m.f. from the microphone and the r.m.s. velocity at an antinode are therefore related in magnitude by a constant factor for sound of fixed frequency. Fields of definite acoustic velocity can be maintained in the tube in terms of the voltage produced at the microphone, and measurements can be made on a series of Rayleigh disks to determine the relations between the dimensions of a Rayleigh disk, the torque and the particle velocity. It should be observed that the construction of such microphones as are suitable for precise measurement is too complicated to allow of direct calculation of velocity at an antinode from the voltage at the microphone. Thus, for example, the resistance and the effective stiffness associated with the diaphragm of a condenser microphone are critically dependent on the motion of the air in the gap between the diaphragm and back plate, and this motion cannot be satisfactorily determined.

---

* This approximation involves an error which is less than 1 % of the pressure for any of the measurements described later in this paper.

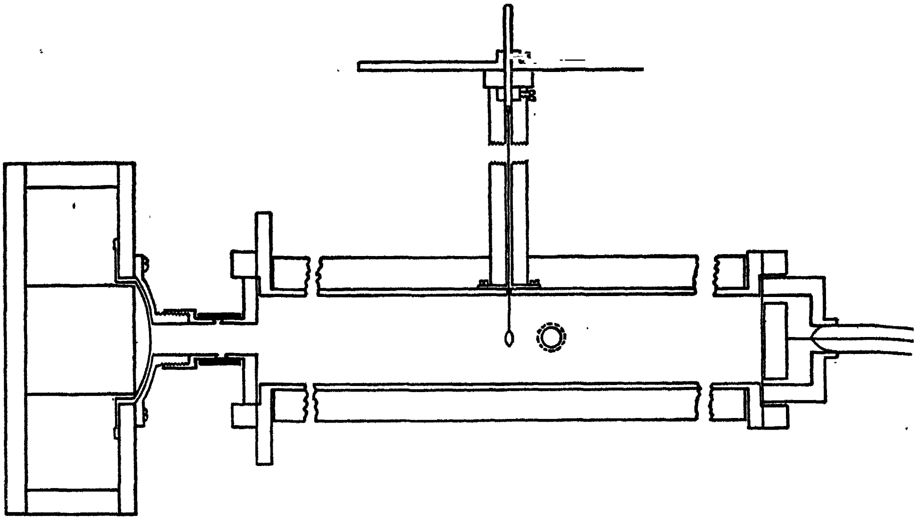


FIGURE 1. Resonant tube and disk suspension.

The tube (figure 1) used in the greater part of the experimental work is 68·4 cm. long and of internal diameter 4·440 cm. The wall of the tube is of brass 0·318 cm. thick and the bore is smooth and closely circular in section. The source of sound consists of a moving-coil loudspeaker unit of the diaphragm type such as is used to drive large exponential horns. This type of driver was chosen because it is capable of transmitting considerable acoustic power to the tube without introducing much harmonic distortion. The unit connects by rubber tube with a hole 1·58 cm. diameter in the centre of a brass plate 0·635 cm. thick which closes the driving end of the tube. The condenser microphone at the other end is fitted with a diaphragm flush with the front of its casing except for a slightly sloping zone at the edge of the diaphragm. The microphone is otherwise of conventional type. Such microphones have been in use over a considerable number of years in this laboratory and have been found to possess very stable pressure sensitivities. The acoustic impedance of the diaphragm is high compared with the characteristic acoustic impedance of the air, and it can be shown that the reflexion at the microphone end of the tube is substantially complete.

A range of Rayleigh disks was constructed to meet the requirements of the experimental work, and details of the disks are listed in table 1. The disks were either ground or turned so as to have edges smooth and square with the faces. The Rayleigh disk in use was supported at the centre of the tube on a fused-silica suspension 36 cm. long and approximately 6μ diameter. The suspension was calibrated by means of circular copper inertia disks which were supported by the suspension and allowed to oscillate in a horizontal plane. The measurement of torsion constant as determined for a suspension from two different inertia disks gave agreement to within about 0·2 %. The suspension was mounted from a torsion head supported by a column attached to the tube and entered the tube through a hole 0·16 cm. square in the wall. The torque on the disk was measured in terms of the twist of the torsion head required to return the disk to a standard position in the field. The direction of the disk relative to the tube was found from the deflexion of a beam of light reflected from the silvered or polished surface of the disk. The light entered the tube through a thin glass window, 0·95 cm. diameter and 0·015 cm. thick, which was recessed in the wall of the tube so as to affect as little as possible the continuity of the inner wall and the area of cross-section of the tube.

TABLE 1. DISKS USED IN THE EXPERIMENTAL WORK

disk	material	mean diam. cm.	maximum variation in diam. from mean cm. × 10^{-4}	thickness cm.	Wood-King correction to torque for lack of inertia of disk %	resonant frequency kcyc./sec. 2 nodal diameters	resonant frequency kcyc./sec. 1 nodal circle
A	glass	0·6284	15	0·0162	0·77	21·6	—
A_1	,,	0·6335	18	0·0162	0·71	21·2	—
B	,,	0·9942	10	0·0142	1·41	10·1	15·9
C	,,	1·0950	15	0·0183	1·18	8·0	12·5
D_1	,,	1·1582	10	0·0211	1·13	8·2	12·9
E	,,	1·9143	20	0·0122	3·22	1·7	2·7
F	,,	2·5959	8	0·0191	2·71	1·5	2·3
D_2	,,	1·1585	10	0·0135	1·69	5·2	8·2
D_3	,,	1·1580	10	0·0343	0·77	13·4	21·0
D_4	,,	1·1595	30	0·0737	0·32	28·4	—
D_5	,,	1·1608	20	0·1219	0·19	47·2	—
D_6	copper	1·1618	56	0·0200	0·34	5·6	8·8
D_7	molybdenum	1·1585	5	0·0064	0·89	2·5	3·9
D_8	,,	1·1570	15	0·0064	0·89	2·5	3·9

The electrical apparatus used in the experimental work needs little description here. The driving unit was supplied with energy from a push-pull oscillator and amplifier capable of providing to the unit an output of about 5 W. The e.m.f. produced at the microphone was amplified by a multi-stage amplifier and passed through a tuned-circuit analyser (vide Churcher & King 1934) to a rectifier moving-coil voltmeter and the overall gain of the apparatus adjusted to give a convenient deflexion of the meter. The input voltage necessary to produce an equal deflexion was then found from an accurately calibrated potentiometer supplied with voltage of the working frequency from the oscillator.

Measurements of frequency were made in terms of a Sullivan valve-maintained tuning fork which had been standardized at the N.P.L. and gave a frequency of 1000·0 cyc./sec. Comparisons of the working frequency and the standard frequency were made with a Cossor double-beam cathode-ray oscilloscope which was connected so that the two sets of plates for vertical deflexion were controlled from the separate supplies. Appropriate adjustment of the time-base makes it possible to obtain a pattern on the screen consisting of two sinusoidal traces moving slowly relatively to each other and containing less, and usually many less, than 30 cycles. Measurement of the slip-rate and of the number of traces in each pattern makes it easy to find the frequency of the oscillator to well within 0·1 % at all the frequencies used in the experimental work.

The entire electrical apparatus is particularly stable, and it is possible to rely on the constancy of indication of the meter over a period of some minutes to as close an accuracy (about 0·2 %) as it is possible to read the meter.

The bulk of the work on the first part of the investigation involves comparison of torque on different disks when placed at the centre of the resonant tube, and little further information is necessary about the precise nature of the sound field beyond the assurance that the disk is in a field of uniform axial velocity. However, the comparison of torque on disks in different tubes (see § 4 (*b*) to follow), depends not only on the uniformity of the field, but also upon the proper location of the antinodes in the tube. With the probe-microphone referred to in § 3 (*a*), explorations of pressure were made along and across the tube and these confirmed that the field conditions within the tube were satisfactory. Further reference to these field measurements will be made only where necessary.

4. RESULTS

In the experimental work described below, the disk was set so that its normal was at 45° with the axis of the tube, except in the case of the determination of variation of torque with angle.

(*a*) *Variation of torque with intensity of sound*

According to König the torque on a Rayleigh disk should vary as the square of the oscillatory velocity and therefore as the square of the voltage from the microphone. The relation, for disk A of table 1, between torque and voltage was found at frequencies of 250, 750, 1250, 4250, and 8250 cyc./sec. over the range of oscillatory velocity 0·13 to 25 cm./sec. (r.m.s.), i.e. intensity levels from 89 to 134 db. or a range of torque of about 37,000 : 1. The graphs in figure 2 show the ratio (torque)/(square of voltage from microphone) plotted against the oscillatory velocity for frequencies of 750 and 4250 cyc./sec.; the observations at the higher intensities (crosses) were made with a thicker suspension. The curve for 750 cyc./sec. shows that the above ratio is constant to within about 1 % in torque over the whole range of intensity. Similar results were obtained at all the frequencies used except at 4250 cyc./sec. In the latter case the graph shows slight curvature, and the departures from constancy in the ratio may be regarded as significant. This small discrepancy is, however, only $2\frac{1}{2}$ % in torque (i.e. a little over 1 % in velocity, if considered in terms of the acoustic field). The substantial constancy of the ratio torque/(voltage)^2 is powerful evidence of (*a*) proportionality between voltage from microphone and acoustic pressure at its diaphragm, and (*b*) proportionality between torque on the disk and the square of the acoustic velocity at the centre of the tube.

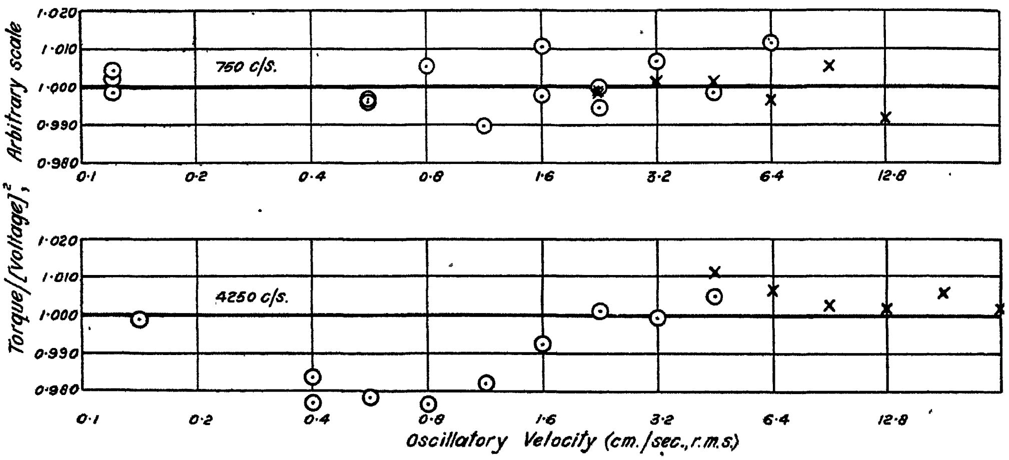


Figure 2. Variation of torque with intensity of sound.

⊙ light suspension. × heavy suspension.

In the remainder of the work the above relations are assumed and the square of the voltage from the microphone has been used to characterize the intensity of the acoustic field at the centre of the resonant tube. In the work described below, it is required to find the manner of variation of the torque with changes of other variables for a disk in a field of fixed intensity; for this purpose, measurements of torque were made for a series of intensities for each experimental condition and the mean of ratios of torque/(voltage)^2 is used for the comparisons.

(*b*) *Effect on torque of proximity of wall of tube*

Comparisons have been made of torque/(voltage)^2 for various disks in each of two tubes. The first tube is that already described and has a diameter of 4·440 cm.; the second tube is also of brass and of similar construction and has a diameter of 12·70 cm. and a length which can be varied up to 300 cm. The latter tube is closed at its ends by heavy pistons into whose faces the microphone and the end-piece from the smaller tube can be fitted. The tube has a central suspension column, and the windows and fittings of each of the tubes are arranged so that the driving unit, the microphone and the disk and suspension can be transferred easily from one tube to the other. The length of the larger tube was adjusted so that the tubes had a common resonant frequency. Observations were made with the aid of the probe microphone to confirm that the acoustic field was uniform and properly located with respect to the centres of the tubes. Measurements were then made of the torque on the disks A, E and F of table 1 in each of the two tubes for frequencies 740 and 1250 cyc./sec. For each disk, the difference between the two values of torque/(voltage)^2 provides a chord of the curve which relates torque/(torque in free space) to the ratio (diameter of disk)/(diameter of tube). From the chords the shape of the whole curve has been synthesized. The sets of data obtained are in reasonable agreement for the two frequencies and a single curve (*a* of figure 3) has been drawn to summarize the results. Attempts to determine the shape of the curve for a frequency of 2750 cyc./sec. were unsatisfactory, since measurements of pressure showed variations across the tube of rather more than 1 %, i.e. equivalent to more than 2 % in torque. Such uncertainty in the field is large enough to swamp the comparatively small effect of the wall of the tube on the torque. At frequencies higher than 2750 cyc./sec. the field variations, and therefore the uncertainties, were larger. Curve *b* in figure 3 shows Tomotika's theoretical results for the two-dimensional analogue of the Rayleigh disk.

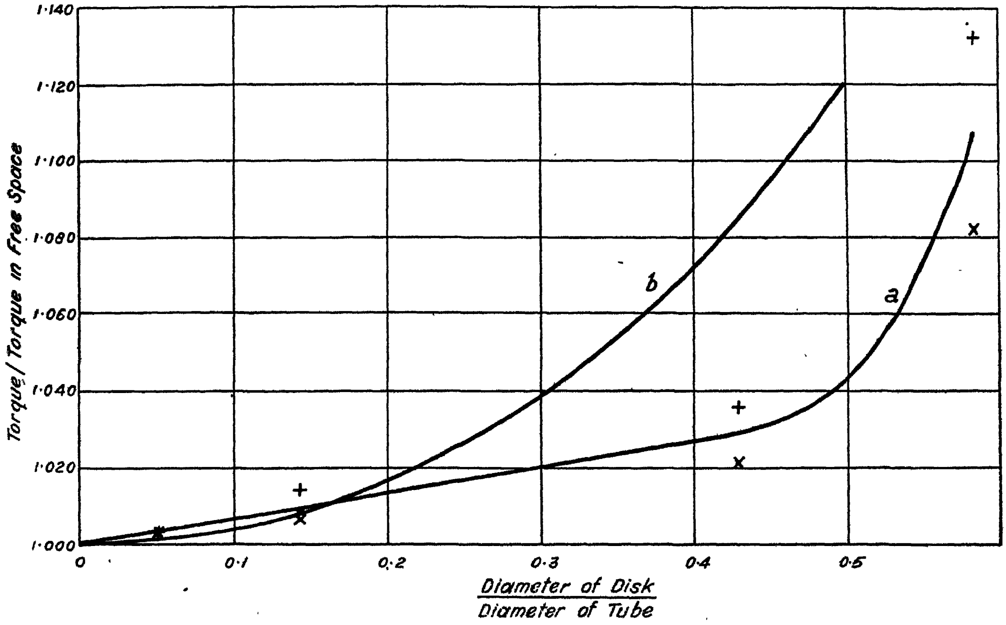


FIGURE 3. Effect on torque of proximity of wall of tube.

(*a*) Experimental curve × 740 cyc./sec. + 1250 cyc./sec.
(*b*) Tomotika's two-dimensional analogue.

(c) *Effect on torque of finite thickness of disk*

A series of disks (D_1–D_8 in table 1), ranging in thickness from 0·0064 to 0·1219 cm., were made of closely similar diameter (1·159 cm.). Measurements were made of the ratio of the torque on each disk to the torque on the disk D, of thickness 0·0211 cm., for the same oscillatory, velocity and the results, suitably corrected with the aid of König's formula for the small variations in diameter, are plotted against the ratio (thickness of disk)/(diameter of disk) in figure 4. The thinner disks were made of copper and molybdenum, since disks of glass would have been so light as to involve large corrections for lack of infinite inertia of the disk. Small corrections for lack of infinite inertia of disk were applied in accordance with King's formula. Agreement between the two sets of results corresponding to 749 and 3272 cyc./sec. is sufficiently good to allow a single curve to be drawn through the points to represent the manner in which the torque on a Rayleigh disk depends upon its thickness. Curves are also included in figure 4 to show the shape of this relation as determined theoretically by König for ellipsoidal disks, and experimentally by Zernov for cylindrical disks.

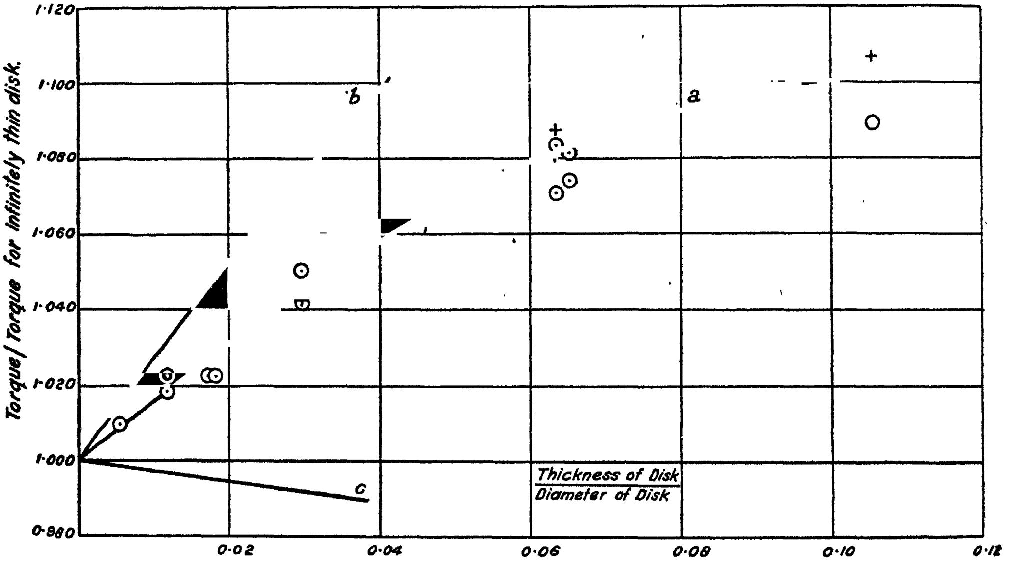


FIGURE 4. Effect on torque of finite thickness of disks.

(a) Author's experimental curve for cylindrical disks.
(b) Zernov's experimental curve for cylindrical disks ⊙ 749 cyc./sec.
(c) König's theoretical curve for ellipsoidal disks + 3272 cyc./sec.

(d) *Variation of torque with radius of disk*

The ratio between torque and radius was investigated for a number of glass disks suspended in the small tube. Comparisons were made in quick succession between torque/(voltage)^2 for disk A and for each of the disks B, C, D and E. The torque was corrected by King's formula for lack of infinite inertia of the disks and by means of figures 3 and 4 of this paper, for proximity of the wall of the tube and for finite thickness of the disk. In figure 5 are plotted the logarithms of the ratios of the torques on the comparison disk and the standard disk against the logarithms of the ratios of the radii of the two disks. The results are shown for the two frequencies 750 and 3270 cyc./sec., and the solid line which is drawn through the set of points has a slope of 3·00. The measurements show that the torque on the disk is closely proportional to the cube of the radius of the disk.

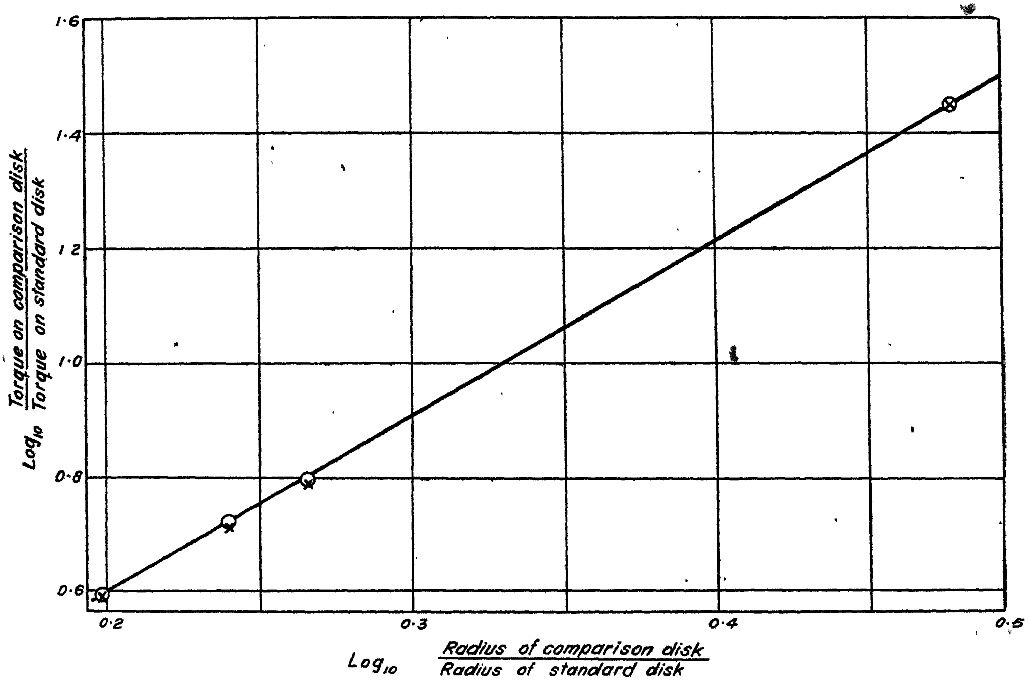


FIGURE 5. Variation of torque with radius of disk.
The solid line has a slope of 3·00. ⊙ 750 cyc./sec. × 3270 cyc./sec.

(*e*) *Variation of torque with angle between disk and stream*

Variations of torque with orientation of the disk relative to the direction of the sound wave were made in the larger tube, since this was fitted with a window subtending a large angle (75°) at the centre of the tube. The direction which the particle velocity in the central section of the tube would have in the absence of the disk was found from the setting of the torsion head for which the disk remained undeflected in either direction, after the sound was turned on in the tube. This direction was found to coincide with that of the geometrical axis of the tube to well within 1°. Measurements of torque/(voltage)^2 were made for various deflected positions of the disk A for frequencies of 740, 1290 and 2900 cyc./sec., and this ratio is compared in table 2 with the sine of twice the angle θ between the normal to the disk and the direction of the stream.

TABLE 2. VARIATION OF TORQUE WITH INCLINATION OF DISK TO STREAM

Frequencies: 740, 1290 and 2900 cyc./sec.

		740 cyc./sec.		1290 cyc./sec.		2900 cyc./sec.	
θ	$\sin 2\theta$	$\frac{L}{E^2}$	$\frac{L}{E^2 \sin 2\theta}$	$\frac{L}{E^2}$	$\frac{L}{E^2 \sin 2\theta}$	$\frac{L}{E^2}$	$\frac{L}{E^2 \sin 2\theta}$
10	0·3420	3312	9684	5945	1738	1162	3397
15	0·5000	4884	9768	8685	1737	1695	3390
20	0·6428	6260	9740	1123	1747	2190	3407
25	0·7660	7486	9774	1343	1754	2614	3413
30	0·8660	8364	9659	1507	1740	2916	3367
35	0·9397	9043	9624	1622	1726	3175	3378
40	0·9848	9437	9583	1727	1754	3337	3389
45	1·0000	9517	9517	1756	1727	3384	3384
48·16	0·9939	—	—	1724	1734	—	—
49·24	0·9891	9297	9398	—	—	—	—

θ is angle between disk and wavefront; L is torque on disk and E is voltage from microphone corresponding to sound giving torque L.

The foregoing series of measurements of torque L at an antinode in a plane-stationary sound field on Rayleigh disks, of widely differing dimensions, has shown that the torque varies closely as the square of the r.m.s. velocity V and as the sine of twice the angle θ. The measurements show that for a very thin disk supported rigidly in free space, the torque would be proportional to the cube of the radius a. Thus the torque which would be exerted on an infinitely thin, immovable Rayleigh disk is given by the expression

$$L = L_1 a^3 V^2 \sin 2\theta. \tag{8}$$

L_1 is here the counterpart of the factor $\frac{4}{3}\rho$ and König's expression (1) and must have the dimensions of a density. L_1 is constant* for air of fixed density. The relation (8) maintains over a wide range of frequency (250–5000 cyc./sec.) and such small discrepancies as may accompany its application lead to errors which, judged by the consistency of the experimental data, are unlikely to be much more than 1 % of the torque.

In the remainder of the investigation the relation between L_1 and König's factor of $\frac{4}{3}\rho$ has been found. For this purpose the determination of L, a and θ present little difficulty and evaluation of the ratio reduces largely to an independent measurement of oscillatory velocity.

Evaluation of the numerical constant involved in the expression for torque

Determinations of sound particle velocity can be carried out without difficulty by a number of methods for frequencies of 50 cyc./sec., and below (vide the work of Zernov and others referred to earlier). At higher frequencies, and especially above

---

* Records were kept of the pressure, temperature and humidity of the air, and suitable small corrections were applied to the torque in proportion to the change in density of the air.

1000 cyc./sec., few methods of measurement exist, and such as have been used are indirect and for the most part difficult to analyse with precision. The method adopted by the present author depends on the determination of particle amplitude from the length of the traces of smoke particles entrained in the sound field. This method has been used by Carrière (1929) and by Andrade (1931), and by Andrade & Parker (1937), and approximates closely to direct measurement of one of the fundamental acoustic quantities, viz. oscillatory amplitude. The validity of the method depends on the extent to which the smoke particles take up the full motion of the medium and this is susceptible to approximate calculation and, to some extent, to experimental demonstration.

The theory of motion of a spherical particle immersed in an oscillating fluid has been worked out by Stokes and others (see Andrade 1931). If w_0 and v_0 are the amplitudes of motion of the sphere and of the surrounding medium (air), if R is the radius of the sphere and ρ_0 is its density, if ν is the kinematic viscosity and ρ the density of the medium and if n is the frequency of the sound:

$$\frac{w_0}{v_0}=\left[\frac{1+3b+\frac{9}{2}b^2+\frac{9}{2}b^3+\frac{9}{4}b^4}{\bar{a}^2+3\bar{a}b+\frac{9}{2}b^2+\frac{9}{2}b^3+\frac{9}{4}b^4}\right]^{\frac{1}{2}} \tag{9}$$

where $\bar{a}=\frac{1}{3}+\frac{2}{3}\frac{\rho_0}{\rho}$, and $b=\frac{1}{R}\sqrt{\frac{\nu}{\pi n}}$. The smoke particles were obtained by igniting 20 mg. of magnesium ribbon in a 2·5 l. flask. Andrade & Parker have estimated the size of MgO particles produced in this general manner and find them to have an effective radius of $0{\cdot}33\mu$. (Observations on smoke from our own apparatus have kindly been made by the staff of the British Cotton Industry Research Association on an R.C.A. electron microscope. The large bulk of the particles which are deposited on the floor of a roughly cubical vessel containing about 20 c.c. of smoke during a short period, about 20 min. after the smoke has been formed, are cubes with edges between $0{\cdot}1$ and $0{\cdot}2\mu$ long.) The ratio w_0/v_0 is calculated according to the formula (9) above for spherical particles of radius $0{\cdot}3\mu$ and the values given in table 3.

TABLE 3. THE SMOKE PARTICLE AS AN INDICATOR OF AIR PARTICLE AMPLITUDE

Calculation on the assumption of a MgO particle of radius $0{\cdot}3\mu$

frequency cyc./sec.	ratio $\frac{\text{amplitude of smoke particle}}{\text{amplitude of air particle}}$
250	0·99996
1250	0·99916
2250	0·9975
3250	0·9947
4250	0·9913
5250	0·9868

It seems reasonable to suppose that this calculation gives at least the right order of magnitude for the ratio w_0/v_0, and that the particles follow the motion of the air to within about 1 % for all frequencies below 4000 cyc./sec. Evidence of the smallness of the discrepancy between the amplitudes of the smoke and the air particles was provided by observing the amplitudes of oscillating MgO smoke particles which had been standing in a still atmosphere for about 20 min. The particles, some of which had become enlarged by coagulation, showed colours which ranged from blue to red and must therefore have varied considerably in size. Nevertheless, no discrepancy between the lengths of the traces of different particles was measurable, even at frequencies of 5000 cyc./sec.

The smoke particle method appears to provide a standard of sound amplitude which is adequately accurate for measurements up to 2000 or 3000 cyc./sec. and is probably not in error by more than 1 or 2 % at frequencies of 5000 cyc./sec.

Apparatus

The resonant tube of 4·440 cm. diameter, described (cf. Andrade & Parker) earlier in the paper, was adapted by the provision of suitable windows and a microscope tube, so that measurements could be made of the amplitude of smoke particle traces at the section of the tube at which the Rayleigh disk was otherwise situated. A cross-section of the tube is shown in figure 6. The particles were illuminated from an arc-lamp (L) and a large-aperture projection lens (P). As a protection against convection currents in the air in the tube, a water-bath (W) was interposed between

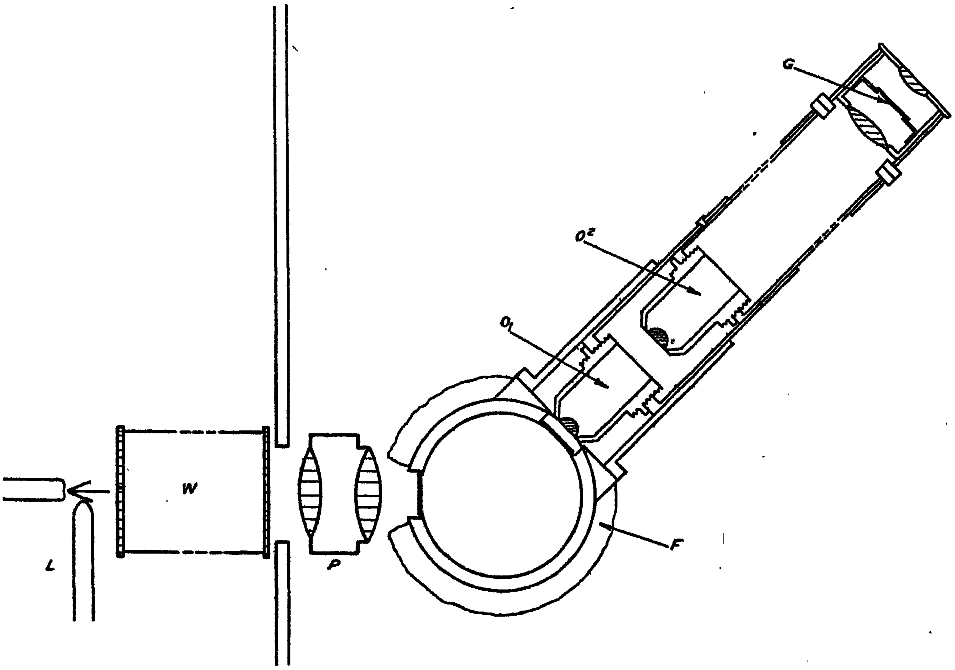


FIGURE 6. Resonant tube fitted for smoke particle amplitude measurement.
L, arc; W, water bath; P, condenser lens; F, felt lagging;
O_1, O_2, microscope objectives; G, eye piece graticule.

the arc and the tube, and the tube was covered with a thick layer of felt (F). A composite microscope objective was constructed from two pieces: and 1 in. Davon (O_1) of n.a. 0·26 and a 16 mm. Beck (O_2) of n.a. 0·28. The composite objective has a focal length of about 2·5 cm., an n.a. of 0·11 and a sufficiently large working distance (about 2·5 cm.) to allow measurements to be made of the particle amplitude at various distances from the axis of the tube. Measurements of the length of traces was made by fitting to the eyepiece a glass graticule consisting of two parallel lines. The intensity of the sound in the tube was adjusted so that the ends of the particle as seen through the microscope appeared just to touch the lines. The effective separation of the lines in the object-space was determined by calibration against an engraved scale which itself had been checked by the N.P.L. A series of graticules was used with spacings which range, when referred to the object-space, from 15·3 to 251·9μ.

Results

At each of the working frequencies, corresponding measurements of length of trace and output voltage from the microphone were obtained and plotted for a

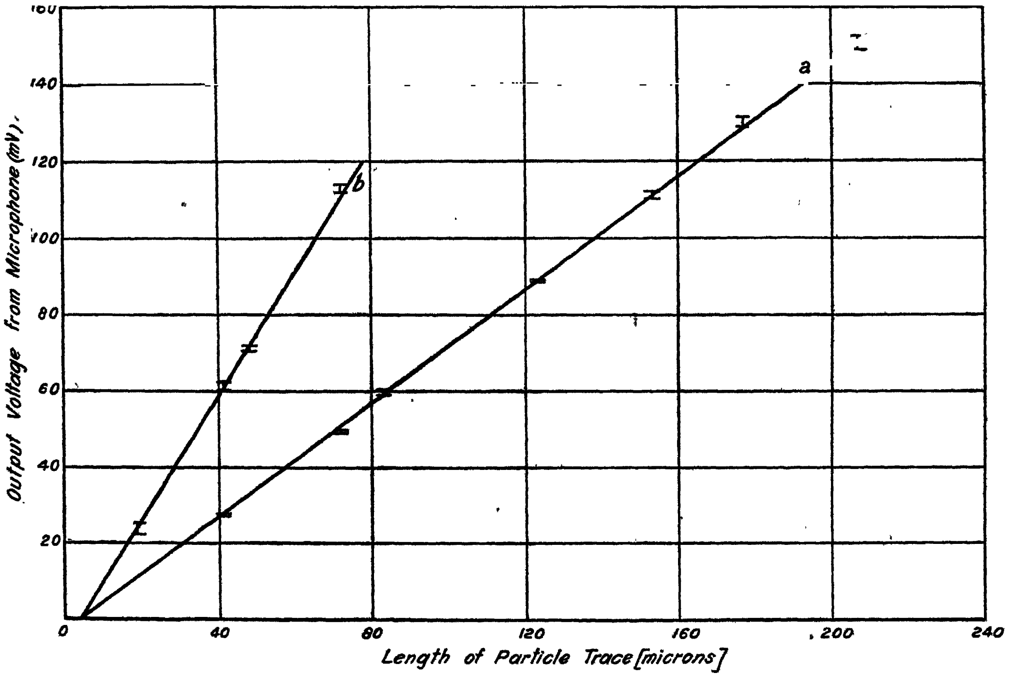


FIGURE 7. Relation between output voltage of microphone and length of smoke particle trace.
(a) 749 cyc./sec. (b) 3275 cyc./sec.

series of about six graticules of different spacing. In each case the traces of about ten separate particles were measured. Figure 7 shows typical curves of apparent length of particle trace against voltage from the microphone for the frequencies

749 and 3275 cyc./sec. In each case the curves are straight lines which cut the length axis at a distance from the origin of between 3 and 4μ. The intercept represents the effective diameter of the diffraction disk produced by the microscope and is in close agreement with the theoretical value (3μ) of the diameter given by $d = k \,.\, 1{\cdot}22\lambda/\text{n.a.}$, where λ is the mean value of the wave-length used ($0{\cdot}55\mu$), and n.a. is the numerical aperture of the objective system (0·11). The factor k, which depends on the visual acuity and has a value of about 0·5, is the ratio of the apparent diameter of the disk to the eye, to the diameter of the first minimum of the Airy diffraction pattern. From the slope of the graph, the value of the ratio (particle amplitude)/(voltage) was obtained. Using this ratio, the velocity at the centre of the tube for any intensity of sound of this frequency could be found in terms of the voltage from the microphone. The tube and associated apparatus were then rearranged and the ratio of torque/(voltage)^2 determined for the Rayleigh disk A. The torque was suitably corrected for the thickness and lack of infinite inertia of the disk, for diffraction and for proximity of the disk to the wall of the tube. The value of L_1 in the formula

$$L = L_1 a^3 v^2 \sin 2\theta$$

was then calculated by combining these data with the measured values of the radius of the disk and of the angle between the disk and the stream. The values of L_1 were found at each of the frequencies: 250, 750, 1250, 1750, 2250, 3250 and 4250 cyc./sec. According to König's formula, the value of L_1 should be $(\frac{4}{3}\rho)$. The ratio of L_1 as determined by experiment, to $(\frac{4}{3}\rho)$ is plotted against frequency in figure 8. In the

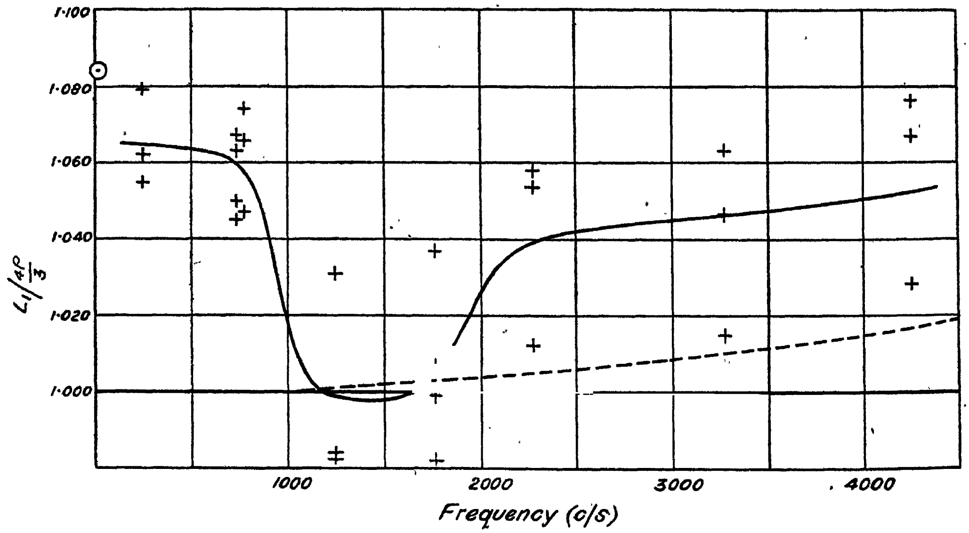


FIGURE 8. Relation between torque calculated from König's formula and torque determined by measurement.

⊙ Merrington and Oatley. + Present author.

Dotted line gives axis of reference on assumption that particle is a sphere of radius $0{\cdot}3\,\mu$ and that formula 9 describes the departure of the smoke particle amplitude from the air particle amplitude.

experimental determination of L_1, the smoke particles have been regarded as following exactly the motion of the air. If we assume that the formula (9) gives the relation between the lengths of the traces of the smoke and the air particles and that the particles have an effective radius of $0{\cdot}3\mu$, the values of $L_1/\frac{4}{3}\rho$ in figure 8 must be reduced in accordance with the results of table 3. The amount of the correction to the ratio is shown by the dotted curve in figure 8 and with the above assumptions this dotted curve should be regarded as the axis to which the ratio $L_1/\frac{4}{3}\rho$ should be referred. For purposes of comparison, the values obtained for the ratio of L_1 to $(\frac{4}{3}\rho)$ at a much lower frequency (9–22 cyc./sec.) by Merrington & Oatley, have been corrected for the effect of finite thickness of the disk by use of figure 4 of this paper and the mean value is plotted as a circled point in figure 8. Application of the correction curve figure 4 improves on the whole the consistency between the individual results of the above authors.

In the region of 750 cyc./sec. a second comparison was made between L_1 and $(\frac{4}{3}\rho)$ using an electromagnetically driven diaphragm as a driving unit for the tube. The ratio agreed with that previously obtained to within 1 %.

5. GENERAL DISCUSSION

In the paper the form of the expression for the torque on a Rayleigh disk of cylindrical shape in a sound field has been investigated over a long range of intensity and frequency. During the investigation, the dependence of the torque on the thickness of the disk and the modification to the torque which results from enclosing the disk in a smooth circular tube have been found. The experiments show that as the disks become very thin and substantially immobile, the expression for the torque tends to a form similar to that put forward on theoretical grounds by König. As the disk becomes thicker the torque tends to increase: for a cylindrical disk of thickness 0·02 of its diameter the torque is increased by 3 %. The thickness therefore has a marked effect on the torque, of a sign which is different from that deduced by König for disks of ellipsoidal form. This conclusion is in agreement with the results of Zernov, who worked with sound of much lower frequency. The effect of the wall of a cylindrical enclosing tube on the torque is shown to be small provided that the diameter of the tube is more than twice that of the disk. Tomotika has obtained theoretically a similar result for the two-dimensional analogue of the Rayleigh disk, using assumptions similar to those of König. The correction to the torque for the tendency of the disk to follow the oscillatory motion of the sound, has been determined independently both by Wood and by King and has been applied to the results of the present work. King's correction—a diffraction correction—for the effect on the torque of the compressibility of the medium has been applied to the results, but it is likely that the correction is small since it was possible to use disks of diameter small compared with the wave-length. Experimental confirmation of the size of this correction is difficult to obtain. Large, and therefore easily measurable, effects are obtained only when the diameter of the disk is of the same order as the wave-length of the sound. The diameter of the enclosing tube must therefore be of the same order as the wave-length and in this case it is difficult to produce longitudinal waves in the air column without introducing transverse waves which distort the field and provide a spurious torque.*

The assumptions of streamline flow, and the absence of discontinuous flow, are made in König's theory of the Rayleigh disk, and the failure of these assumptions would invalidate König's expression for the torque. It must be borne in mind that the agreement in general form between the experimental results and those predicted by König's expression is not in itself evidence of the applicability of König's theory. If we commence with the premise that the torque must depend upon a simple product of some power each of the density of the air, the radius of the disk and the velocity of the air, it can be shown by the method of dimensions that the torque varies directly as the density, directly as the cube of the radius and directly as the square of the particle velocity. It is therefore probable that even if turbulence were present in the sound field near the disk, an expression of the form

$$L = L_1 a^3 V^2 \sin 2\theta$$

would continue to describe the torque on the disk. Justification of the validity of König's analysis is best afforded by comparison of the experimental value of L_1 with the König value ($\frac{4}{3}\rho$) of this factor. The comparison has been made and is described in the second part of this paper.

The comparison of L_1 with $\frac{4}{3}\rho$ requires the independent determination of particle velocity, and this has been made by means of measurement of amplitude of smoke particles in the sound field. The comparisons show that the torque exerted on the disk is larger than is given by König's formula by an amount between 0 and 7 %. The magnitude of this discrepancy varies with frequency and is least in the neighbourhood of 1200 cyc./sec. At 250 cyc./sec. the discrepancy is about 7 % of the torque, and this agrees within 2 % with the observations of Merrington & Oatley when working under very different experimental conditions at frequencies between 9 and 22 cyc./sec. With regard to the cause of this lack of agreement between the experimental and the König values for the torque, there is little which can be added save in confirmation of the observations of Merrington & Oatley. Investigation of the motion of smoke-laden air in the sound field in the neighbourhood of the Rayleigh disk confirms the existence, over the whole range of frequency and intensity covered by the present work, of Andrade vortices. There is strong evidence of the continued existence of these circulations for the lowest velocities at which it is possible to estimate movement of the particles. No abrupt transition to vortex-free conditions could be observed. The apparent absence of transition is supported by the linearity of the relation between torque on the disk and the square of voltage from the microphone. The slight departure from linearity which is shown in figure 2 for 4250 cyc./sec. does not suggest that the torque on the disk, even at this frequency, tends to be consistently lower for very small velocities.

---

* The corrections which have been discussed above can in practice be made sufficiently small (say, less than 2 % of the torque) not to introduce serious errors in the application of König's expression. Thus for a glass disk 0·63 cm. diameter and 0·007 cm. thick, the correction for diffraction (in air) is 1 % of the torque for 3800 cyc./sec. and the correction due to proximity of the wall of an enclosing tube 5 cm. diameter is −0·6 %. The correction for the thickness of the disk is −1·8 % and that for the mobility of the disk is +1·8 % of the torque. The last two of these corrections can each be reduced by using a slightly thinner disk made from a dense material such for example as polished stainless steel.

At 750 cyc./sec. an attempt was made to estimate the magnitude of the torque on the disk due to the steady circulations. The circulations were made visible by blowing magnesium oxide smoke into the air near the disk, and the shapes of the principal sections of the pattern were photographed using flat-beam illumination. Measurements were made of the velocity, for a given intensity level of the sound, at various points in the field and the magnitude of the streaming velocity close to the disk was estimated from the data. From the velocity an assessment was made in the following manner of the torque which should accompany the circulations. The maximum value of the defect of pressure near the surface of the disk was calculated by Bernoulli's law from the maximum velocity near the disk, and this defect of pressure was multiplied by one-third of the area of the disk to set an upper limit to the force on each face of the disk. The force was then multiplied by the diameter of the disk. These operations might be expected to give an overestimate of the magnitude of the torque due to the vortices. Investigation showed that this torque, although of correct sign, appears to be rather less than 1 % of the total torque and therefore insufficient to account for the departure from König's formula.

The discrepancies of figure 8 are too large to be accountable in terms of experimental uncertainty. Such residual errors as may be associated with the uncertainty in the experimental measurements described in this paper are unlikely to be much larger than 1 % in particle velocity, i.e. 2 % in torque. The measurements of torque, radius of disk, density, frequency and voltage were made with a precision sufficient to determine L_1 to within 0·2 %. The largest uncertainties are due to three possible causes. Harmonic distortion of the acoustical waveform might have introduced errors in L_1 of up to 0·5 % during measurements of torque or of particle amplitude. In the higher frequency region, i.e. above 3000 cyc./sec., the observed non-uniformity of the sound field might have introduced errors of 1 % in the determination of L_1. Measurements of smoke particle amplitude could be made to about 1 %, that is, corresponding to an uncertainty of about 2 % in L_1.

The investigation confirms the stability of performance of the Rayleigh disk and justifies its continued use as a standard of reference of acoustical intensity. Calculations of the particle velocity of the medium in terms of the torque on the disk must take account of the failure of the numerical factor $\frac{4}{3}$ in König's formula and suitable corrections amounting from 0 to 3·5 % of the particle velocity must be applied according to the results of figure 8. For the large bulk of measurements made in acoustics with microphones, an accuracy of 1 db. in intensity is adequate; in such work the corrections enumerated above, which amount to no more than 0·3 db., are barely significant.

The author is indebted to Mr B. G. Churcher and to Mr A. J. King for authorization of the work in this paper and for their continued interest and encouragement during its progress. The author wishes to thank the staff of the British Cotton Industry Research Association for their aid in determining with their electron microscope the size of the smoke particles. The author expresses his thanks to Dr A. P. M. Fleming, C.B.E., Director of Metropolitan-Vickers Electrical Co. Ltd., for permission to publish this paper.

REFERENCES

Andrade 1931 *Proc. Roy. Soc.* A, **134**, 445.
Andrade & Parker 1937 *Proc. Roy. Soc.* A, **159**, 507.
Barnes & West 1927 *J. Instn Elect. Engrs*, **65**, 871.
Carrière 1929 *J. Phys. Radium*, **10**, 198.
Churcher & King 1934 *J. Instn Elect. Engrs*, **75**, 411.
King 1935 *Proc. Roy. Soc.* A, **153**, 17.
König 1891 *Ann. Phys., Lpz.* **43**, 43.
Merrington & Oatley 1939 *Proc. Roy. Soc.* A, **171**, 505.
Rayleigh 1880 *Proc. Camb. Phil. Soc.* **4**, 18.
Rayleigh 1882 *Phil. Mag.* **14**, 186.
Tomotika 1937 *Proc. Phys. Math. Soc. Japan*, **19**, 328.
Wood 1935 *Proc. Phys. Soc.* **47**, 779.
Zernov 1908 *Ann. Phys., Lpz.* **26**, 79.

A note on the Hamiltonian theory of quantization

BY T. S. CHANG, PH.D., *Central University, Chungking*

(*Communicated by P. A. M. Dirac, F.R.S.—Received* 11 *July* 1944)

In this note a study is made of the field equations which are obtained from varying a Lagrangian subject to auxiliary conditions. It is shown that with proper choice of conjugate variables and the Hamiltonian, they can be brought to canonical forms and thus permit quantization in the usual way. It is then pointed out that for fields with some conjugate variables missing, it is sometimes possible to introduce such auxiliary conditions without affecting the field equations but with the result that all the new canonical variables are present, thus allowing the application of the standard methods of quantization. To illustrate this, the Maxwell field is quantized subject to $\nabla . A + c^{-1}\, \partial\phi/\partial t = 0$. The usual supplementary condition on ψ, $(\nabla . A + c^{-1}\, \partial\phi/\partial t)\,\psi = 0$, is found to be shifted to other conditions on ψ. Though in the beginning there are apparently some complications, the final result is simple.

The Hamiltonian theory of a continuum and its quantization was first given by Heisenberg & Pauli (1929) and was later developed in another form by Weiss (1936, 1938*a*, 1938*b*). The latter form has the advantage that it is throughout relativistically invariant, and thus no explicit proof of the relativistic invariance of the equations obtained at various stages is necessary.

According to these authors, the field equations are the Euler equations obtained by taking the minimum of the integral of a certain invariant function L (known as

Lagrangian) over the space formed by the independent variables. In this note we shall take the minimum after imposing auxiliary conditions on the dependent variables and see the alterations which must be made on the canonical equations and the subsequent quantization. In view of the possibility that a general field theory may contain such auxiliary conditions, such a study is perhaps not entirely meaningless.

Sections 1, 2 and 3 give the general theory of quantization for fields with such auxiliary conditions. The method and the notations employed by Weiss will be followed. It will be seen that the modifications due to the presence of auxiliary conditions are few. To apply the theory, we shall quantize the Maxwell field after imposing on it the condition

$$\nabla . A+\frac{1}{c}\frac{\partial \phi}{\partial t}=0.$$

Though in the beginning there are more complications than in the usual theories, the final result is simple.

1. EULER'S EQUATIONS AND THE COMPLETE VARIATION

Let z^α be the dependent variables ($\alpha = 1, 2, \ldots, \nu$), x^i be the independent variables ($i = 1, 2, \ldots, n$), and z^α_i be the derivatives $\mathfrak{d}z^\alpha/\mathfrak{d}x^i$.† Take the minimum of

$$I=\int_G^{(n)} \mathscr{L}(x^i, z^\alpha, z^\alpha_i)\,dx \quad \left(dx=\prod_{i=1}^{n} dx^i\right), \tag{1}$$

subject to the f conditions

$$f^{(\xi)}(x^i, z^\alpha, z^\alpha_i)=0 \quad (\xi=1,2,\ldots,f). \tag{2}$$

Using Lagrange's method of unknown multipliers, it is found that

$$\frac{\partial \mathscr{L}}{\partial z^\alpha}+\Sigma\mu_\xi\frac{\partial f^\xi}{\partial z^\alpha}-\frac{\mathfrak{d}}{\mathfrak{d}x^i}\left(\frac{\partial \mathscr{L}}{\partial z^\alpha_i}+\Sigma\mu_\xi\frac{\partial f^\xi}{\partial z^\alpha_i}\right)=0, \tag{3}$$

which together with (2) gives z^α as functions of x^i corresponding to a minimum of I when the values of x and z on the boundary S of the region G are kept fixed. (2) and (3) are the field equations.

Now find the complete variation of I. Using Lagrange's method of unknown multipliers, then

$$\delta I=\int_G^{(n)}\left\{\frac{\partial \mathscr{L}}{\partial z^\alpha}+\Sigma\eta_\xi\frac{\partial f^\xi}{\partial z^\alpha}-\frac{\mathfrak{d}}{\mathfrak{d}x^i}\left(\frac{\partial \mathscr{L}}{\partial z^\alpha_i}+\Sigma\eta_\xi\frac{\partial \mathscr{L}}{\partial z^\alpha_i}\right)\right\}\delta z^\alpha\,dx+\int_S^{(n-1)}(P_\alpha\,\delta z^\alpha+X_i\,\delta x^i)\,du,\ddagger \tag{4}$$

---

† Differentiation with respect to one of the n variables x^i is denoted by $\mathfrak{d}/\mathfrak{d}x^i$, and differentiation with respect to one of the $n+\nu+n\nu$ variables z^α, x^i, z^α_i by $\partial/\partial z^\alpha$, $\partial/\partial x^i$, $\partial/\partial z^\alpha_i$.

‡ The two δz^α in (4) have different meanings. No attempt will be made to explain them here, the reader being referred to Weiss's paper (1936).

where u^r ($r = 1, 2, \ldots, n-1$) are the parameters for the surface S and η_ξ are the multipliers which may yet be arbitrary functions of x^i,

$$P_\alpha = N_i\left\{\frac{\partial \mathscr{L}}{\partial z^\alpha_i} + \Sigma\eta_\xi \frac{\partial f^\xi}{\partial z^\alpha_i}\right\}, \tag{5}$$

$$X_i = N_l\left\{L\delta^l_i - z^\alpha_i\left[\frac{\partial \mathscr{L}}{\partial z^\alpha_l} + \Sigma\eta_\xi \frac{\partial f^\xi}{\partial z^\alpha_l}\right]\right\}, \tag{6}$$

N_l being defined by

$$a^l N_l = \begin{vmatrix} a^1 & \cdots & a^n \\ \dfrac{\mathfrak{d}x^1}{\mathfrak{d}u^1} & & \dfrac{\mathfrak{d}x^n}{\mathfrak{d}u^1} \\ \vdots & & \vdots \\ \dfrac{\mathfrak{d}x^1}{\mathfrak{d}u^{n-1}} & & \dfrac{\mathfrak{d}x^n}{\mathfrak{d}u^{n-1}} \end{vmatrix}. \tag{7}$$

Taking δI as due to variations of z^α from a solution of the field equations (2) and (3), we may let the η's be the μ's in (3), and then

$$\delta I = \int_S (P^*_\alpha \delta z^\alpha + X^*_i \delta x^i)\, du, \tag{8}$$

where P^*_α and X^*_i are given by the right-hand sides of (5) and (6) with η_ξ replaced by μ_ξ.

The conservation equations are modified. Letting

$$U^l_i = L\delta^l_i - z^\alpha_i\left[\frac{\partial \mathscr{L}}{\partial z^\alpha_l} + \Sigma\mu_\xi \frac{\partial f^\xi}{\partial z^\alpha_l}\right], \tag{9}$$

so that $X^*_i = N_l U^l_i$, then

$$\frac{\mathfrak{d}U^l_i}{\mathfrak{d}x^l} = \frac{\partial L}{\partial x^i} + \Sigma\mu_\xi \frac{\partial f^\xi}{\partial x^i}. \tag{10}$$

2. HAMILTONIAN FORMULATION

Introduce a co-ordinate u^0, so that

$$N_l(\mathfrak{d}x^l/\mathfrak{d}u^0) = 1. \tag{11}$$

This co-ordinate is denoted by w in Weiss's papers, and is, for symmetry in notation, changed to u^0 in the present note. It is seen at once that

$$N_i(\partial L/\partial z^\alpha_i) = \partial L/\partial \mathfrak{z}^\alpha_0 \tag{12}$$

and

$$X_i \delta x^i = (L - \mathfrak{z}^\alpha_0 P_\alpha)\,\delta u^0 - \mathfrak{z}^\alpha_r P_\alpha \delta u^r,\dagger \tag{13}$$

where $\mathfrak{z}^\alpha_r$ denotes $\mathfrak{d}z^\alpha/\mathfrak{d}u^r$ and $\mathfrak{z}^\alpha_0$ denotes $\mathfrak{d}z^\alpha/\mathfrak{d}u^0$. For shortness, the function $L - \mathfrak{z}^\alpha_0 P_\alpha$ is denoted by H'.

It is necessary to note that P_α is defined by (5) with quantities η_ξ whose nature has not been specified. Thus H' is a function of x^i (or u^r, u^0), z^α, z^α_i and η_ξ. With the

---

† r runs always from 1 to $n-1$.

help of (2) and (5) it may be transformed into a function of x^i, z^α, $\tilde{z}^\alpha_r$ and P_α, say H. Evidently dH must be dH' plus a linear combination of df^ξ and

$$d\left(P_\alpha - \frac{\partial L}{\partial \tilde{z}^\alpha_0} - \Sigma\eta_\xi \frac{\partial f^\xi}{\partial \tilde{z}^\alpha_0}\right),$$

the linear combination being so chosen that its sum with dH' does not contain $d\tilde{z}^\alpha_0$ and $d\eta_\xi$. Noting that

$$dH' + \Sigma\eta_\xi\left(\frac{\partial f^\xi}{\partial x^i}dx^i + \frac{\partial f^\xi}{\partial z^\alpha}dz^\alpha + \frac{\partial f^\xi}{\partial z^\alpha_i}dz^\alpha_i\right)$$

can be reduced down to

$$\frac{\partial L}{\partial u^0}du^0 + \frac{\partial L}{\partial u^r}du^r + \Sigma\eta_\xi\frac{\partial f^\xi}{\partial x^i}dx^i - \tilde{z}^\alpha_0 dP_\alpha + \left(\frac{\partial L}{\partial z^\alpha} + \Sigma\eta_\xi\frac{\partial f^\xi}{\partial z^\alpha}\right)dz^\alpha + \left(\frac{\partial L}{\partial \tilde{z}^\alpha_r} + \Sigma\eta_\xi\frac{\partial f^\xi}{\partial \tilde{z}^\alpha_r}\right)d\tilde{z}^\alpha_r, \tag{14}$$

and thus does not contain $d\tilde{z}^\alpha_0$ and $d\eta_\xi$, it may be concluded that it is dH. Needless to say, η_ξ in (14) are functions of x^i, z^α, $\tilde{z}^\alpha_r$ and P_α determined by (2) and (5).

From (14), it follows that

$$\frac{\partial H}{\partial P_\alpha} + \frac{\mathfrak{d}z^\alpha}{\mathfrak{d}u^0} = 0. \tag{15}$$

It is easy to see now that the equation

$$\frac{\partial H}{\partial z^\alpha} - \frac{\mathfrak{d}}{\mathfrak{d}u^r}\frac{\partial H}{\partial \tilde{z}^\alpha_r} - \frac{\mathfrak{d}}{\mathfrak{d}u^0}P_\alpha = 0 \tag{16}$$

is precisely (3) with η_ξ in place of μ_ξ. Hence the set of equations (2), (5), (16) is completely equivalent to the set of equations (2), (3). Equation (15) follows from (2) and (5) and the definition of the function H. Hence we have the set of equations (2), (5), (15) and (16) to solve for z^α. But if (15) and (16) are independent, which they must be in a general case (since there is no reason to the contrary), they are already sufficient. Hence (2) and (5) must be redundant equations, i.e. they are deducible from (15) and (16). A rigorous proof of this statement will not be attempted here.

The foregoing says that the set of equations (15), (16) is equivalent to the set (2), (3). It says also that if P_α and z^α satisfy (15), (16), η_ξ are simply the μ_ξ's in (3), and no distinction between the present P_α and the P^*_α in § 1 is necessary. From the appearance of the equations (15), (16), H may be called the Hamiltonian, (15), (16) the canonical equations, and P_α, z^α the conjugate variables.

Conversely, given a H and the equations (15), (16), of which the solution is known to satisfy (2), it is possible to find a Lagrangian whose minimum subject to (2) gives the given field equations. In fact, a possible Lagrangian is obtained from $H + \tilde{z}^\alpha_0 P_\alpha$ by eliminating P_α with the help of (15).

The conservation equation is, from (14), (15), (16),

$$\frac{\mathfrak{d}}{\mathfrak{d}u^0}\int H du = \int \frac{\partial H}{\partial u^0}du, \quad \frac{\partial H}{\partial u^0} = \frac{\partial L}{\partial u^0} + \Sigma\eta_\xi\frac{\partial f^\xi}{\partial u^0}, \tag{17}$$

η_ξ being functions of x^i, z^α, $\tilde{z}^\alpha_r$, P_α obtained by solving for them from (2) and (5).

3. Passage to quantum theory

Let us see if the theory of relative integral invariants as given by Weiss (1938*a*) can be taken over here. According to him, the space of the independent variables in consideration should be of a cylindrical type, by which is meant that if the independent variables are x, y, z and the time t, the space is bounded by two space-like surfaces S_1 and S_2 and a time-like surface.[†] Let us consider the proper specification of data on S_1 which are here called the initial data and those on the time-like surface which are called the boundary data. From the form of equations (15), (16), it is at once seen that values of z^α and P_α on S_1 form a complete set of initial data and those of z^α on the time-like surface a complete set of boundary data. This may be explained by the fact that though on S_1 values may be assigned to only $2\nu - f$ of the z^α and $\tilde{z}_0^\alpha$'s, since there are f relations between them, yet the μ's in (3) are arbitrary, and strictly speaking their values at S_1 form a part of the initial data. This allows arbitrary values to be given to P_α^* and z^α on S_1, or P_α and z^α on S_1 (since P_α^* and P_α are identical so long as P_α, z^α satisfy (15), (16)), and thus to treat P_α, z^α as a complete set of initial data.

With the initial data P_α^*, z^α, the boundary data z^α and the formula (8), it is not difficult to see that all brackets $[P_\alpha^*(u), P_\beta^*(u')]$, $[P_\alpha^*(u), z^\beta(u')]$, $[z^\alpha(u), z^\beta(u')]$,[‡] as defined by Weiss (1938*a*), are integral invariants. Thus

$$[P_\alpha, P_\beta] = [z^\alpha, z^\beta] = 0, \quad [P_\alpha(u), z^\beta(u')] = \delta_\alpha^\beta \delta(u-u'), \tag{18}$$

and thus

$$\left[f(u), \int -P_\alpha(u')\,\tilde{z}_r^\alpha(u')\,du'\right] = \frac{\mathfrak{d}f(u)}{\mathfrak{d}u^r}. \tag{19}$$

As usual, the canonical equations give

$$\frac{\partial f(u)}{\partial u^0} + \left[f(u), \int H(u')\,du'\right] = \frac{\mathfrak{d}f(u)}{\mathfrak{d}u^0}. \tag{20}$$

In quantum mechanics we define

$$[a,b]_- = ab - ba, \quad [a,b]_+ = ab + ba. \tag{21}$$

Assuming the law

$$\epsilon\left[f(u), \int H(u')\,du'\right]_- + \frac{\partial f(u)}{\partial u^0} = \frac{\mathfrak{d}f(u)}{\mathfrak{d}u^0}, \tag{22}$$

where ϵ is a constant depending on the choice of u^0, $[z, P]_-$ and $[z, P]_+$, etc., do not change with u^0 if they are initially c-numbers. (i) For the case $[z, z]_-$, $[z, P]_-$, etc., are c-numbers, the particular choice

$$[P_\alpha, P_\beta]_- = [z^\alpha, z^\beta]_- = 0, \quad \epsilon[P_\alpha(u), z^\beta(u')]_- = \delta_\alpha^\beta \delta(u-u'), \tag{18·1}$$

etc., gives field equations identical with the classical ones. Equation (19) is replaced by

$$\epsilon\left[f(u), \int -P_\alpha(u')\,\tilde{z}_r^\alpha(u')\,du'\right]_- + \frac{\partial f(u)}{\partial u^r} = \frac{\mathfrak{d}f(u)}{\mathfrak{d}u^r}, \tag{19·1}$$

[†] A space-like surface (or a time-like surface) is one in which the vector joining a point in the surface to another point in its immediate neighbourhood also lying in the surface is necessarily a space-like (or may be a time-like) vector.

[‡] u, u' stand for $u^1, \ldots, u^{n-1}$ and $u'^1, u'^2, \ldots, u'^{n-1}$.

which completes with (22) to form a relativistically invariant equation. (ii) For the case $[z,z]_+$, $[z,P]_+$, etc., are c-numbers, the particular choice

$$[P_\alpha, P_\beta]_+ = [z^\alpha, z^\beta]_+ = 0, \quad \epsilon[P_\alpha(u), z^\beta(u')]_+ = l\delta^\beta_\alpha(u-u'), \tag{18·2}$$

etc., with l as an arbitrary real or complex number gives also classical field equations, provided that $H = \Sigma \ldots z^{\cdot\cdot}_{\cdot\cdot}P_{\cdot\cdot}$, and on passing to quantum mechanics, we let H be

$$H = (2l)^{-1}\Sigma \ldots [(l+1)\,z^{\cdot\cdot}_{\cdot\cdot}P_{\cdot\cdot} + (l-1)\,P_{\cdot\cdot}z^{\cdot\cdot}_{\cdot\cdot}]. \tag{23}$$

So far, it is difficult to find other simple forms of H, so that (18·2) and (22) give the classical field equations. The equation (19) is now changed to

$$\left(\frac{\epsilon}{2l}\right)\left[f(u), -\int\{(l+1)\,\tilde{z}^\alpha_r P_\alpha + (l-1)\,P_\alpha \tilde{z}^\alpha_r\}(u')\,du'\right]_- + \frac{\partial f(u)}{\partial u^r} = \frac{\mathfrak{d} f(u)}{\mathfrak{d} u^r}. \tag{19·2}$$

Thus if we have classically $\Sigma \ldots z^{\cdot\cdot}_{\cdot\cdot}P_{\cdot\cdot}$, it may seem that there are many Fermi-Dirac schemes of quantization, corresponding to different values of l. But they are actually equivalent, for through a proper transformation of the z and P, one scheme may change into another.

The question which arises if one requires z, P to be permanently real (or imaginary) can be almost always solved if L is chosen properly. For (i) let us restrict ourselves for definiteness to the case in which there are no auxiliary conditions, the Lagrangian L is symbolically quadratic in z and none of the P's are missing. Let u^0, u^r be all real and let us assume ϵ to be purely imaginary. Now suppose we choose L so that if z^α, z^α_i are real c-numbers, so is L. On letting z and P be all real dynamical variables (or all imaginary) at one u^0, which is consistent with (18·1), it is found that H at this u^0 is real, and thus z and P remain real (or imaginary). Case (ii) with its special form of the Hamiltonian can be realized only in cases in which L is symbolically linear in the dynamical variables z^α and also in the variables y^β $(\alpha, \beta = 1, 2, \ldots, \nu)$ and the conjugate variables of y do not exist, or in equivalent cases. Assuming that all the conjugate variables of z^α, say $P(z^\alpha)$, exist and are independent, $H(z^\alpha, z^\alpha_r, P(z^\alpha))$ may be introduced as

$$L - \tilde{z}^\alpha_0 P(z^\alpha),$$

whereupon the Euler equations for z^α and y^β become precisely the canonical equations with the Hamiltonian H and the conjugate variables z^α, $P(z^\alpha)$. With real u^0, u^r and imaginary ϵ as before and with l chosen to be 1 for definiteness, we choose L so that if z^α and $P(z^\alpha)$ are real c-numbers, so is H. Then on letting z be real and P be imaginary (or vice versa) at one u^0, which is consistent with (18·2), we find that H is real at u^0 and thus the reality of z, etc., remains.

This section will be concluded with a word on the passage from operator equations of the type

$$\frac{\mathfrak{d}}{\mathfrak{d} u^0} f(z, P, u^0, u^r) = \frac{\partial f}{\partial u^0} + \epsilon\left[f, \int H(u')\,du'\right]_- \tag{24}$$

to a Schrödinger equation, concerning which no clarification, though mostly a repetition of well-known facts, can be considered as superfluous. For simplicity, assume $n = 1$, $u^0 = t$, $\nu = 1$, i.e. there is only one z and one P, both being functions of t only. It is necessary to note that H in (24) is a dynamical variable at time t. Hence in order to get (24), it is necessary to assume for the wave function ψ at time t, say ψ_t, the equation

$$i\hbar\frac{d\psi_t}{dt} = H(z_t, P_t, t)\,\psi_t, \tag{25}$$

where, in order not to forget that z, P are functions of t, a suffix t is attached to them. Let us choose the fixed ψ in the Heisenberg picture to coincide with the moving ψ in the Schrödinger picture at $t = 0$. Thus the fixed z, P in the Schrödinger picture, say z_s, P_s, coincide with the moving z, P in the Heisenberg picture at $t = 0$, say, z_0, P_0. Thus the correct procedure to get a Schrödinger equation is to solve z_t, P_t in terms of z_0, P_0, t, to substitute into (25), to replace z_0, P_0 by z_s, P_s, and to insert $\Sigma\psi_{z_s'}(z_s'\,|\,t)$ for ψ_t. In case H does not contain t explicitly, one may use the fact that

$$H(z_t, P_t) = H(z_0, P_0),$$

and the task of solving for z_t, P_t in terms of z_0, P_0, t and substituting into $H(z_t, P_t)$ can be avoided.

Two notes must be added. First, from the appearance of (25), one might be tempted to put ψ_t as

$$\Sigma\psi_{z_t'}(z_t'\,|\,t),$$

where $\psi_{z'_t}$ is the eigenstate for z_t with the eigenvalues z_t'. Using the convention for employing the primes so that z_t' takes the same numerical value for all t, it is possible to find $d\psi_{z_t'}/dt$ from the law of change of z_t, and with that it is possible to get from (25) an equation giving $d(z_t'\,|\,t)/dt$ in terms of $(z_t'\,|\,t)$. It is true that this equation is usually simple, but we have the disadvantage that $\Sigma(t\,|\,z_t')\,z_t'(z_t'\,|\,t)$ is not the expected value of z at time t. To get the desired value $\phi_0 z_t \psi_0$ or $\phi_t z_0 \psi_t$ with the help of $(z_t'\,|\,t)$, it is necessary to know $(z_t'\,|\,z_0'')$, which cannot be obtained unless the equation of motion for z_t is integrated. Secondly, it is easy to note that the procedure of replacing $H(t, z_t, P_t)$ by $H(t, z_s, P_s)$, as is actually done by current practices is wrong. Consider the various derivatives to t of

$$H(t, z_t, P_t) - H(t, z_0, P_0)$$

taken at time $t = 0$. They are zero only when, at $t = 0$,

$$\left[\frac{\partial H}{\partial t}, H\right]_- = 0, \quad \left[\frac{\partial^2 H}{\partial t^2}, H\right]_- = 0,$$

etc. Since this is in general not so, $H(t, z_t, P_t)$ and $H(t, z_s, P_s)$ are not equal.

4. APPLICATION TO ELECTRODYNAMICS

The foregoing theory of quantization for fields whose field equations are deduced from varying a Lagrangian subject to auxiliary conditions can sometimes be applied with advantage to fields whose field equations are deduced from varying a Lagrangian under no auxiliary conditions but with some of the P's missing. Let the latter field equations admit the integrals $f^{\xi}(x^i, z^{\alpha}, z^{\alpha}_i) = 0$. Then, evidently, these field equations can be deduced from varying the same Lagrangian subject to $f^{\xi} = 0$. Thus the methods in §§ 1, 2, 3 may be applied. Generally speaking, it may be found that in so doing all the new P's are present. If this is so, the quantization may proceed in a straightforward way.

To illustrate this, let us quantize the Maxwell field using the auxiliary condition

$$\nabla . A + \frac{\partial \phi}{\partial t} = 0.\dagger \tag{26}$$

This is different from the procedure adopted by Fermi (1932) and discussed by Dirac (1935); they assume that

$$\left(\nabla . A + \frac{\partial \phi}{\partial t}\right) \psi = 0 \tag{27}$$

holds for ψ representing a state that can actually occur in nature. Here (26) is an operator equation and thus holds for all ψ.

Let $$ds^2 = g_{\rho\theta}\, du^{\rho}\, du^{\theta}$$

(ρ, θ, ... run from 0 to 3, and r, s, ... from 1 to 3), and assume for simplicity that the g's are constants. Hence leaving out the Lagrangian which involves the matter only, we have

$$L = \frac{1}{8\pi}\left(\frac{\partial z^{\rho}}{\partial u^{\theta}}\frac{\partial z_{\rho}}{\partial u^{\theta}} - \frac{\partial z^{\rho}}{\partial u^{\theta}}\frac{\partial z^{\theta}}{\partial u^{\rho}}\right) - j_{\theta} z^{\theta}, \ddagger \tag{28}$$

where $\partial j^{\theta}/\partial u^{\theta} = 0$. A straightforward application of the methods of the above sections gives

$$P_r = \frac{1}{4\pi}\left(\frac{\partial z_r}{\partial u_0} - \frac{\partial z^0}{\partial u^r}\right), \quad P_0 = \eta, \tag{29}$$

and $$H = \frac{1}{8\pi}\left(\frac{\partial z^r}{\partial u^s}\frac{\partial z_r}{\partial u_s} - \frac{\partial z^r}{\partial u^s}\frac{\partial z^s}{\partial u^r}\right) - j_{\theta} z^{\theta} + P^0_0 \frac{\partial z^r}{\partial u^r} - P^0_r \frac{\partial z_0}{\partial u_r} - 2\pi P^0_r P^r_0, \tag{30}$$

where P^0_r, P^0_0 are written in place of P_r, P_0 in order to show its transformation character. The canonical equations are

$$\left.\begin{aligned} \frac{\partial z^r}{\partial u^r} + \frac{\partial z^0}{\partial u^0} &= 0, \\ -4\pi P^r_0 - \frac{\partial z_0}{\partial u_r} + \frac{\partial z^r}{\partial u^0} &= 0, \\ -j_0 + \frac{\partial}{\partial u^r} P^r_0 - \frac{\partial}{\partial u^0} P^0_0 &= 0, \\ -j_r - \frac{1}{4\pi}\frac{\partial}{\partial u^s}\left(\frac{\partial z_r}{\partial u_s} - \frac{\partial z^s}{\partial u^r}\right) - \frac{\partial}{\partial u^0} P^0_r - \frac{\partial}{\partial u^r} P^0_0 &= 0, \end{aligned}\right\} \tag{31}$$

---

† The velocity of light c is taken to be unity for convenience.

‡ In this section j_{θ} are treated as c-numbers.

of which the last three, apart from terms containing P^0_0, are precisely Maxwell's equations. As can be seen, the auxiliary equation appears in the canonical equation for $\mathfrak{d}z^0/\mathfrak{d}u^0$.

To get down to Maxwell's equations, it is sufficient to assume that for ψ occurring in nature,

$$P^0_0\psi = 0, \quad \left(-j_0+\frac{\partial}{\partial u^r}P^r_0\right)\psi = 0 \tag{32}$$

at $u^0 = 0$. This is of the same nature as the supplementary condition (27) mentioned in the beginning of this section. It is not difficult to see from (31), which holds for any ψ, that if ψ satisfies (32) at a certain u^0, it does so for all subsequent u^0. Thus we have the first two equations of (31) and

$$\left(-j_0+\frac{\partial}{\partial u^r}P^r_0\right)\psi = 0,$$

$$\left\{-j_r-\frac{1}{4\pi}\frac{\partial}{\partial u^s}\left(\frac{\partial z_r}{\partial u_s}-\frac{\partial z^s}{\partial u^r}\right)-\frac{\partial}{\partial u^0}P^0_r\right\}\psi = 0,$$

for all u^0. In the following, we shall assume the commutation relations

$$[P,P]_- = [z,z]_- = 0, \quad \epsilon[P^0_\rho(u), z^\theta(u')]_- = \delta^\theta_\rho\,\delta(u-u'), \tag{33}$$

etc.; hence the operator equations $P^0_0 = 0$ and $-j_0+(\partial/\partial u^r)\,P^r_0 = 0$ at $u^0 = 0$ cannot be assumed on account of their incompatibility with (33).

Let us employ the co-ordinates $u^\rho = u_\rho = (it, x, y, z)$ so that $g_{\rho\theta} = \delta_{\rho\theta}$ and the ϵ in equation (24) is equal to $-1/\hbar$. On putting

$$z_\theta = \Sigma z_{\theta,\mathbf{k}}\,e^{i\mathbf{k}\mathbf{x}}, \quad P_\theta = \Sigma P_{\theta,\mathbf{k}}\,e^{i\mathbf{k}\mathbf{x}}, \quad j_\theta = \Sigma j_{\theta,\mathbf{k}}\,e^{i\mathbf{k}\mathbf{x}},$$
$$\mathbf{k} = (k_x, k_y, k_z), \quad \mathbf{x} = (x, y, z),$$

the assumed commutation relations become

$$[P_{\rho,-\mathbf{k}'}, z_{\theta,\mathbf{k}}]_- = -\hbar\delta_{\rho\theta}\,\delta_{\mathbf{k}\mathbf{k}'}, \tag{34}$$

etc. To proceed further, choose the axes for each Fourier wave so that $\mathbf{k}$ becomes $(k, 0, 0)$, k being $+\sqrt{(\Sigma k_r^2)}$. Let us perform a transformation

$$\left.\begin{aligned}
z_{0,\mathbf{k}} &= e^{ikt}[-if^+_{1,-\mathbf{k}}-f^+_{0,-\mathbf{k}}(\tfrac{1}{2}kt+\tfrac{1}{4}i)]+e^{-ikt}[if^-_{0,\mathbf{k}}-f^-_{1,\mathbf{k}}(\tfrac{1}{2}kt-\tfrac{1}{4}i)],\\
z_{1,\mathbf{k}} &= e^{ikt}[f^+_{1,-\mathbf{k}}-if^+_{0,-\mathbf{k}}(\tfrac{1}{2}kt-\tfrac{1}{4}i)]+e^{-ikt}[f^-_{0,\mathbf{k}}+if^-_{1,\mathbf{k}}(\tfrac{1}{2}kt+\tfrac{1}{4}i)],\\
z_{2,\mathbf{k}} &= e^{ikt}f^+_{2,-\mathbf{k}}+e^{-ikt}f^-_{2,\mathbf{k}},\\
z_{3,\mathbf{k}} &= e^{ikt}f^+_{3,-\mathbf{k}}+e^{-ikt}f^-_{3,\mathbf{k}},\\
(4\pi/k)\,P_{0,-\mathbf{k}} &= -if^+_{0,\mathbf{k}}\,e^{ikt}-if^-_{1,-\mathbf{k}}\,e^{-ikt},\\
(4\pi/k)\,P_{1,-\mathbf{k}} &= f^+_{0,\mathbf{k}}\,e^{ikt}-f^-_{1,-\mathbf{k}}\,e^{-ikt},\\
(4\pi/k)\,P_{2,-\mathbf{k}} &= f^+_{2,\mathbf{k}}\,e^{ikt}-f^-_{2,-\mathbf{k}}\,e^{-ikt},\\
(4\pi/k)\,P_{3,-\mathbf{k}} &= f^+_{3,\mathbf{k}}\,e^{ikt}-f^-_{3,-\mathbf{k}}\,e^{-ikt},
\end{aligned}\right\} \tag{35}$$

where $P_{1,-\mathbf{k}}$ denotes the component of $P_{-\mathbf{k}}$ along the direction of $\mathbf{k}$, not that of $-\mathbf{k}$. The f are therefore dynamical variables involving t explicitly. The relations (35) are so chosen that for the case $j^{\theta} = 0$, the f's are constants of motion. In terms of f, the commutation relations become

$$[f^{+}, f^{+}]_{-} = [f^{-}, f^{-}]_{-} = 0, \quad [f^{+}_{\rho,\mathbf{k}}, f^{-}_{\theta\mathbf{k}'}]_{-} = -\frac{2\pi\hbar}{k}\delta_{\rho\theta}\delta_{\mathbf{k}\mathbf{k}'}. \tag{36}$$

The reality condition of the field is given by

$$f^{+}_{1\mathbf{k}} = -\overline{f^{-}_{0\mathbf{k}}}, \quad f^{+}_{0\mathbf{k}} = -\overline{f^{-}_{1\mathbf{k}}}, \qquad f^{+}_{2\mathbf{k}} = \overline{f^{-}_{2\mathbf{k}}}, \quad f^{+}_{3\mathbf{k}} = \overline{f^{-}_{3\mathbf{k}}}, \tag{37}$$

with $\bar{v}$ denoting the adjoint of v. The integral of H over x, y, z becomes

$$H_{\text{rad.}} + H_{\text{I}},$$

$$H_{\text{I}} = -\Sigma j_{\theta,-\mathbf{k}} z_{\theta,\mathbf{k}},$$

$$H_{\text{rad.}} = \Sigma\frac{k^2}{4\pi}\{f^{-}_{2\mathbf{k}}f^{+}_{2\mathbf{k}} + f^{+}_{2\mathbf{k}}f^{-}_{2\mathbf{k}} + f^{-}_{3\mathbf{k}}f^{+}_{3\mathbf{k}} + f^{+}_{3\mathbf{k}}f^{-}_{3\mathbf{k}} + 2f^{+}_{0\mathbf{k}}f^{-}_{0\mathbf{k}} + 2f^{-}_{1\mathbf{k}}f^{+}_{1\mathbf{k}} - f^{-}_{1\mathbf{k}}f^{+}_{0\mathbf{k}}\}, \tag{38}$$

and the initial conditions on ψ become the equations

$$\left\{f^{+}_{0,\mathbf{k}} - \frac{2\pi}{k^2}ij_{0,-\mathbf{k}}e^{-ikt}\right\}\psi = 0, \quad \left\{f^{-}_{1,\mathbf{k}} + \frac{2\pi}{k^2}ij_{0,\mathbf{k}}e^{ikt}\right\}\psi = 0, \tag{39}$$

for $t = 0$. According to the above, if we assume (39) to hold at $t = 0$, they hold for all t.

A further transformation

$$\left.\begin{aligned} b_{0\mathbf{k}} &= \frac{i}{\sqrt{2}}(f^{+}_{0\mathbf{k}} + f^{-}_{1\mathbf{k}}), & a_{0\mathbf{k}} &= \frac{1}{\sqrt{2}}(f^{-}_{0\mathbf{k}} - f^{+}_{1\mathbf{k}}), \\ b_{1\mathbf{k}} &= \frac{1}{\sqrt{2}}(f^{+}_{0\mathbf{k}} - f^{-}_{1\mathbf{k}}), & a_{1\mathbf{k}} &= \frac{i}{\sqrt{2}}(f^{-}_{0\mathbf{k}} + f^{+}_{1\mathbf{k}}), \\ b_{2\mathbf{k}} &= \frac{1}{\sqrt{2}}(f^{+}_{2\mathbf{k}} + f^{-}_{2\mathbf{k}}), & a_{2\mathbf{k}} &= -\frac{i}{\sqrt{2}}(f^{+}_{2\mathbf{k}} - f^{-}_{2\mathbf{k}}), \\ b_{3\mathbf{k}} &= \frac{1}{\sqrt{2}}(f^{+}_{3\mathbf{k}} + f^{-}_{3\mathbf{k}}), & a_{3\mathbf{k}} &= -\frac{i}{\sqrt{2}}(f^{+}_{3\mathbf{k}} - f^{-}_{3\mathbf{k}}), \end{aligned}\right\} \tag{40}$$

gives real a and b satisfying

$$[a, a]_{-} = [b, b]_{-} = 0, \quad [b_{\rho\mathbf{k}}, a_{\theta\mathbf{k}}]_{-} = -\frac{2\pi\hbar i}{k}\delta_{\rho\theta}\delta_{\mathbf{k}\mathbf{k}}. \tag{41}$$

$H_{\text{rad.}}$ becomes

$$\Sigma\frac{k^2}{4\pi}\{b_{2\mathbf{k}}b_{2\mathbf{k}} + a_{2\mathbf{k}}a_{2\mathbf{k}} + b_{3\mathbf{k}}b_{3\mathbf{k}} + a_{3\mathbf{k}}a_{3\mathbf{k}} + 2(b_{1\mathbf{k}}a_{0\mathbf{k}} - b_{0\mathbf{k}}a_{1\mathbf{k}}) + \tfrac{1}{2}(b_{0\mathbf{k}}b_{0\mathbf{k}} + b_{1\mathbf{k}}b_{1\mathbf{k}})\}, \tag{42}$$

and the initial condition on ψ

$$(b_{0\mathbf{k}} - \gamma_{0\mathbf{k}})\psi = (b_{1\mathbf{k}} - \gamma_{1\mathbf{k}})\psi = 0 \tag{43}$$

for $t = 0$, where

$$\gamma_{0\mathbf{k}} = -\frac{\sqrt{(2)}\,\pi}{k^2}(j_{0,-\mathbf{k}}e^{-ikt} - j_{0,\mathbf{k}}e^{ikt}),$$

$$\gamma_{1\mathbf{k}} = \frac{\sqrt{(2)}\,\pi i}{k^2}(j_{0,-\mathbf{k}}e^{-ikt} + j_{0,\mathbf{k}}e^{ikt}).$$

5. CONTINUATION

Now pass to the Schrödinger equation. According to the discussion in § 3, it runs as

$$i\hbar\frac{d\psi_t}{dt} = H(t)\,\psi_t,[†] \quad H = H_{\text{rad.}}+H_{\text{I}}+H_{\text{M}}, \tag{44}$$

where H_{M} is the Hamiltonian for matter. Let α be the dynamical variables for matter. Let the dynamical variables α, z, P, etc., at $t = 0$ coincide with the respective fixed variables α_s, z_s, P_s in the Schrödinger picture, and let $H(\alpha, z, P)$ not contain t explicitly, so that $H(\alpha_t, z_t, P_t)$ may be replaced by $H(\alpha_s, z_s, p_s)$. The equation (44) becomes

$$i\hbar\frac{\partial}{\partial t}(z_s'\alpha_s' \mid t) = \Sigma(z_s'\alpha_s' \mid H \mid z_s''\alpha_s'')\,(z_s''\alpha_s'' \mid t). \tag{45}$$

It may be noted that the variables z_s, P_s may also be transformed to f_s or a_s, b_s, the transformation law being (35), (40) with t put as zero. This we shall do, obtaining $H(\alpha_s, a_s, b_s)$, which involves no t explicitly. In the following, all notations concerned with dynamical variables refer to fixed variables in the Schrödinger picture (unless otherwise mentioned), and the suffix s will be dropped.

The representative $(\alpha', n'_{2\mathbf{k}}\dots, n'_{3\mathbf{k}}\dots, b'_{0\mathbf{k}}\dots, b'_{1\mathbf{k}}\dots \mid -)$ will be used, where

$$n_{2\mathbf{k}} = \frac{k}{4\pi\hbar}(b_{2\mathbf{k}}b_{2\mathbf{k}}+a_{2\mathbf{k}}a_{2\mathbf{k}})-\tfrac{1}{2},$$

$$n_{3\mathbf{k}} = \frac{k}{4\pi\hbar}(b_{3\mathbf{k}}b_{3\mathbf{k}}+a_{3\mathbf{k}}a_{3\mathbf{k}})-\tfrac{1}{2},$$

with

$$b_{2\mathbf{k}} = \left(\frac{\pi\hbar}{k}\right)^{\frac{1}{2}}(n_{2\mathbf{k}}{}^{\frac{1}{2}}e^{iw_{2\mathbf{k}}}+e^{-iw_{2\mathbf{k}}}n_{2\mathbf{k}}{}^{\frac{1}{2}}),$$

$$a_{2\mathbf{k}} = -\left(\frac{\pi\hbar}{k}\right)^{\frac{1}{2}}i(n_{2\mathbf{k}}{}^{\frac{1}{2}}e^{iw_{2\mathbf{k}}}-e^{-iw_{2\mathbf{k}}}n_{2\mathbf{k}}{}^{\frac{1}{2}}),$$

$$a_{0\mathbf{k}} = \frac{2\pi\hbar i}{k}\frac{\partial}{\partial b_{0\mathbf{k}}}, \quad a_{1\mathbf{k}} = \frac{2\pi\hbar i}{k}\frac{\partial}{\partial b_{1\mathbf{k}}}, \quad \text{etc.}$$

(e^{iw}, e^{-iw} are well-known matrices.) Substitution of the above into (45) gives a rather lengthy equation, but it may be simplified as follows.

After assuming (43) at $t = 0$, these equations follow for all t. Thus

$$(b_{0\mathbf{k}}-\gamma_{0\mathbf{k}})_t\,\psi = 0, \quad (b_{1\mathbf{k}}-\gamma_{1\mathbf{k}})_t\,\psi = 0 \tag{46}$$

for all t in a Heisenberg picture, or

$$(b_{0\mathbf{k}}-\gamma_{0\mathbf{k}})_{t=0}\,\psi_t = 0, \quad (b_{1\mathbf{k}}-\gamma_{1\mathbf{k}})_{t=0}\,\psi_t = 0, \tag{47}$$

[†] It is hoped that no confusion arises from using H to denote what should be denoted by $\int H\,dx\,dy\,dz$.

in a Schrödinger picture. On taking a representation in which all $j_{0\mathbf{k}}$ are diagonal,† we find

$$(\alpha', n' \ldots, b' \ldots \mid t) = (\alpha', n'_{2\mathbf{k}} \ldots, n'_{3\mathbf{k}} \ldots \mid t)$$
$$\times \prod_{\mathbf{k}} \delta^{(\frac{1}{2})}\left(b'_{0\mathbf{k}} + \frac{\sqrt{2\pi}}{k^2}[j'_{0,-\mathbf{k}} - j'_{0,\mathbf{k}}]\right) \prod_{\mathbf{k}} \delta^{\frac{1}{2}}\left(b'_{1\mathbf{k}} - \frac{\sqrt{2\pi i}}{k^2}[j'_{0,-\mathbf{k}} + j'_{0,\mathbf{k}}]\right). \ddagger \quad (48)$$

Substituting this into the Schrödinger equation, multiplying by

$$\pi\delta^{(\frac{1}{2})}(b'_0 + \ldots)\, \pi\delta^{(\frac{1}{2})}(b'_1 - \ldots).$$

and integrating over all $b'_{0\mathbf{k}}$ and $b'_{1\mathbf{k}}$, all the terms in H containing a_0 or a_1 are dropped out. The simplified equation becomes

$$i\hbar \frac{\partial}{\partial t}(\alpha', n'_{2\mathbf{k}} \ldots, n'_{3\mathbf{k}} \ldots \mid t) = \sum_{\alpha''} (\alpha' \mid H \mid \alpha'')(\alpha'', n' \ldots \mid t)$$
$$+ \sum_{\mathbf{k}} k\hbar(n'_{2\mathbf{k}} + n'_{3\mathbf{k}})(\alpha' n' \ldots \mid t) - \sum_{\mathbf{k}} \frac{2\pi}{k^2} j'_{0,\mathbf{k}} j'_{0,-\mathbf{k}}(\alpha', n' \ldots \mid t) - \sum_{\alpha'', n''} \sum_{2,3} \sum_{\mathbf{k}} \left(\frac{2\pi\hbar}{k}\right)^{\frac{1}{2}}$$
$$\times (\alpha', n' \ldots \mid n_{2\mathbf{k}}^{\frac{1}{2}} e^{iw_{2\mathbf{k}}} j_{2,\mathbf{k}} + e^{-iw_{2\mathbf{k}}} n_{2\mathbf{k}}^{\frac{1}{2}} j_{2,-\mathbf{k}} \mid \alpha'', n'' \ldots)(\alpha'', n'' \ldots \mid t). \quad (49)$$

The third term in the right-hand side is precisely the Coulomb energy times the representative $(\alpha' n' \mid t)$.

From (48), it is not difficult to see that the probability of having α taking the values α', n taking n', is proportional to $\mid (\alpha' n' \mid t) \mid^2$, and the average value of n to $\Sigma n' \mid (\alpha' n' \mid t) \mid^2$. If $(\beta' z' P' \mid t)$ is required, where β is any other dynamical variable of matter, we have only to apply the ordinary transformation theory. All the information required can be derived from $(\beta' z' P' \mid t)$.

The present quantization amounts to shifting the supplementary condition from (27) to equations (32) for $t = 0$. It has the advantage of having P_0 and the commutation relations (33) which do not place z_0, P_0 and the other z, P on different footings. It enables one to calculate the complete wave function $(\alpha' n' b' \mid t)$ in a not too complicated manner. But as can be seen, both from the general character of the theory and from the equation (49), the usual divergencies in the field theory still remain. Hence, practically speaking, the improvement brought forth by the present theory, if any, is small.

The present quantization can be applied to Proca's field. There we see that from the field equations we may obtain $\partial z^\theta / \partial u^\theta = 0$, and thus we may introduce it in the very beginning as an auxiliary condition. In the subsequent development, the departures from the above calculations are of a trivial kind and will therefore be omitted.

† This can be assumed if j_0 at different $\mathbf{x}$ and the same t commute, which holds generally.

‡ $\delta^{\frac{1}{2}}(x)$ is defined by $\delta^{\frac{1}{2}}(x) = 0$ if $x \neq 0$, and $\int_{-\infty}^{+\infty} [\delta^{\frac{1}{2}}(x)]^2 = 1$. It is introduced to make $\int \mid (\alpha', n'b' \mid -) \mid^2 db'$ finite. Evidently it can be written as $\{\delta(x)\}^{\frac{1}{2}}$ without causing confusion. Though in ordinary mathematical analysis the usage of such functions warrants special care, they can be used rather freely in quantum mechanics.

In conclusion, I wish to thank my former teacher, Professor R. H. Fowler, for his kind encouragement.

REFERENCES

Dirac 1935 *Quantum mechanics*, 2nd ed., last chapter.
Fermi 1932 *Rev. Mod. Phys.* **4**, 125.
Heisenberg & Pauli 1929 *Z. Phys.* **56**, 1.
Weiss 1936 *Proc. Roy. Soc.* A, **156**, 192.
Weiss 1938*a* *Proc. Roy. Soc.* A, **169**, 102.
Weiss 1938*b* *Proc. Roy. Soc.* A, **169**, 119.

The occurrence and properties of molecular vibrations with $V(x) = ax^4$

BY R. P. BELL

(*Communicated by C. N. Hinshelwood, F.R.S.—Received* 16 *November* 1943)

It is shown that in certain modes of vibration of plane rings the potential energy for small displacements is proportional to the fourth power of the displacement, provided that there is free rotation about the bonds of the ring. This type of vibration is termed a 'fourth-power vibration'. It is likely to occur in cyclobutane and its derivatives, in a number of halides having the formula X_2Y_6, and in the hydrides of group III elements.

The energies and wave functions of the first four levels of a one-dimensional oscillator with $V(x) = ax^4$ have been derived by a method of successive approximations, and asymptotic formulae are given for the higher levels. The wave functions are qualitatively similar to those of a harmonic oscillator, but the energy levels differ considerably. A comparison is made between energy levels for oscillators with $V(x) = a_q|x^q|$ and different values of q.

The selection rule for dipole radiation from a fourth-power vibration is discussed. Overtones will be more numerous than in the spectrum of a harmonic oscillator. Estimates are made of the spectrum frequencies of fourth-power vibrations in actual molecules, with special reference to cyclobutane and diborane. For these two molecules there are observed infra-red frequencies of approximately the expected value. The isotope effect should provide a means of discriminating experimentally between harmonic and fourth-power vibrations.

The contribution of a fourth-power vibration to any thermodynamic function will differ from that of a harmonic vibration with the same fundamental spectrum frequency. Figures are given for the specific heat, where the difference should be detectable experimentally. In the general case $V(x) = a_q|x^q|$ the energy levels derived from the quantum theory lead to expressions for the thermodynamic functions which agree with the predictions of classical theory at high temperatures.

INTRODUCTION

The energy levels and wave functions of a one-dimensional oscillator have been derived for a number of different potential energy functions, the simplest of which is the harmonic oscillator with $V(x) = ax^2$. All these functions resemble the harmonic oscillator in that the plot of $V(x)$ against x has a finite radius of curvature at the equilibrium point $x = 0$, i.e. if $V(x)$ is expanded in powers of x near $x = 0$, the coefficient of x^2 is not zero. Similarly, in the normal co-ordinate treatment of molecular vibrations it is always assumed in the first instance that the potential energy is a quadratic function of the displacements, further terms being introduced only as corrections to the quadratic form. It will be shown in the present paper that there may be some types of molecular vibration for which these methods of treatment are inadequate, since for small displacements the potential energy is proportional to the *fourth* power of the displacement. In these cases if $V(x)$ is expanded in powers of x near $x = 0$, the coefficient of x^2 is zero, and the plot of $V(x)$ against x has an infinite radius of curvature at $x = 0$. It is therefore of interest to investigate the properties of a one-dimensional oscillator in which $V(x) = ax^4$, and a vibration of this type will be referred to as a *fourth-power vibration.*

VIBRATIONS OF PLANE RINGS

The kind of vibration described above may occur in molecules containing a plane ring of four or more atoms, which may have other atoms or groups attached to them. This can be illustrated by considering the ring $ABCD$ in figure 1. In one of its modes of vibration the atoms will move initially at right angles to the plane of the ring, as shown by the arrows. It is easily seen from the geometry of the system that a first-order infinitesimal displacement x produces only a second-order change in the distances between any pair of atoms, or in the angles between any pair of bonds attached to the same atom. If therefore the forces in the ring can be represented by either a valency-force or a central-force system, the expression for the potential energy will contain no term in x^2, and if the usual normal co-ordinate treatment is carried out one frequency will turn out to be zero.

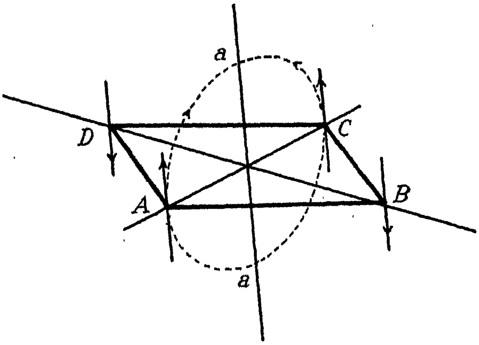
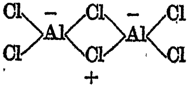


FIGURE 1

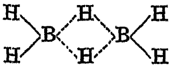
A somewhat similar position is met with for the out-of-plane bending vibration of the planar molecule XY_3, which is assigned a zero frequency by the ordinary valency-force or central-force treatment. However, in this case, although there is no first-order change in single valency-angles, there is a first-order distortion of the disposition of valencies round the atom X, which may be expressed in terms of the angle between the bond XY and the plane Y_3. It is therefore not surprising that for boron trichloride the force constant for the out-of-plane bending is approximately the same as for bending in the plane of the molecule (Anderson, Lassettre & Yost 1936). On the other hand, in the vibration represented in figure 1 the valencies attached to any given atom suffer no first-order displacement of any kind relative to one another, and this remains true also when there are other atoms or groups attached to any of the atoms $ABCD$.

The nature of the forces which resist distortion from a plane ring-configuration will vary from one case to another, and will be related to the path followed by the atoms in a finite displacement. Two extreme cases may be distinguished. In the first case, if the forces between unlinked atoms can be neglected, then the distances between adjacent (linked) atoms will remain unchanged, the atoms will describe circles (e.g. a, a in figure 1), and the restoring force will be due to distortions of valency angles. In the second case, if the diagonal distances AC and BD are fixed by strong forces, the atoms will continue to move in straight lines, and there will be a change both in the valency angles and in the bond distances AB, BC, CD and DA. In either case it is clear from the geometry of the system that for a small displacement the changes in distances and angles are all proportional to the square of the displacement, and hence that the change in potential energy is proportional to the fourth power of the displacement. In an actual system the motion would represent a compromise between these two extreme cases, resembling the first case more closely, since the forces between linked atoms will normally be greater than those between unlinked ones. The potential energy will also follow a fourth-power law in this intermediate case, and the constant a in $V(x) = ax^4$ can be expressed in terms of the ordinary quadratic force constants of a central-force or valency-force system.

It has been tacitly assumed so far that there is no restriction on the relative directions of bonds attached to *different* atoms, i.e. that there is free rotation about the bonds in the ring. If rotation is not free there may be a quadratic term in the potential energy even when there are no groups attached to the ring atoms: for example, in the motion represented in figure 1 the bonds AD and BC are twisting about AB. We can therefore exclude from consideration here plane rings containing bonds of higher order than single bonds (e.g. benzene), since in these there is undoubtedly restricted rotation about the ring bonds. In the cycloparaffin series rings containing six or more atoms are non-planar, so that the present treatment will apply directly only to cyclobutane and its derivative, and possibly also to cyclopentane, though there is some evidence (Aston, Fink & Schumann 1943) that the latter may not be planar. There are also many compounds of formula X_2Y_6 which contain a planar four-ring of single bonds: thus dimeric aluminium chloride can be written



and there is direct evidence that this configuration is correct for the halides of aluminium, indium and gallium (Palmer & Elliott 1938; Brode 1940). The same structure probably applies to other dimeric halides such as Fe_2Cl_6 and Au_2Cl_6, and it also occurs in more complex co-ordinated halogen compounds (Mann 1938). Moreover, it has been recently shown (Longuet-Higgins & Bell 1943) that the hydrides B_2H_6 and Ga_2H_6 probably have the bridged structure



and that the same type of bridge linking occurs in many of their derivatives. There are thus many examples of planar four-rings containing no multiple bonds, which are of interest in the present context.

There is probably always some resistance to rotation even about single bonds, which may introduce an energy term proportional to x^2 in the type of vibration considered above. Experimental and theoretical knowledge on this question is at present very incomplete, but the available data suggest that hindered rotation is unlikely to be the decisive factor in the present problem, though it is likely to produce a correction to the energy expression $V(x) = ax^4$ for relevant values of the displacement. In any case a study of the ring vibrations should give information about the actual form of the potential-energy curve, and hence about the extent to which rotation is restricted. It should be noted that the existence of an appreciable hindrance to rotation will in general modify the normal strain-free equilibrium configuration of the ring, e.g. cyclobutane would not be exactly planar. However, electron diffraction measurements on methylenecyclobutane (Bauer & Beach 1942) indicate a planar configuration for this compound, and the vibrational spectrum of cyclobutane itself (Wilson 1943) does not admit of a non-planar structure: hence it is probable that resistance to rotation is small in this compound.

Energy levels and wave functions

If the potential energy is given by $V(x) = ax^4$, the wave equation is

$$\frac{d^2\psi}{dx^2}+\frac{8\pi^2 m}{h^2}(W-ax^4)\psi = 0, \tag{1}$$

which on making the substitutions

$$\xi = \left\{\frac{8\pi^2 ma}{h^2}\right\}^{\frac{1}{6}} x, \quad \lambda = 4W\left\{\frac{\pi^4 m^2}{h^4 a}\right\}^{\frac{1}{3}}, \tag{2}$$

becomes

$$\frac{d^2\psi}{d\xi^2}+(\lambda-\xi^4)\psi = 0. \tag{3}$$

Equation (3) is not soluble by known functions or in series, and the general equations given by Dunham (1932*b*) for energy levels are not applicable in the absence of a term in ξ^2. The numerical difference-equation method of Kimball & Shortley (1934) gives both eigen-values and eigen-functions, and we have applied it to the first four levels of equation (3), using a difference $\Delta\xi = 0{\cdot}2$. Starting with a trial function of roughly the right form, satisfactory convergence is obtained after 4–6 approximations. The eigen-values of λ thus obtained are given in column (*a*) of table 1. They have an estimated accuracy of about 2 %. Figure 2 shows the first four wave functions obtained by the method given in table 1, which are qualitatively similar to those for a harmonic oscillator. The difference is illustrated by the broken curve, which is the wave function of the lowest state of a harmonic oscillator having the same energy as the lowest state of (1).

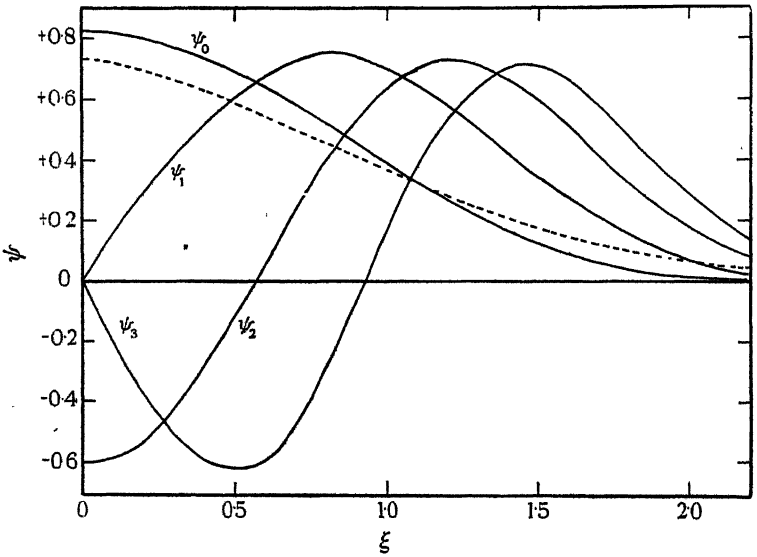


FIGURE 2

TABLE 1. EIGEN-VALUES OF EQUATION (3)

	(*a*)	(*b*)	(*c*)	(*d*)
λ_0	1·09	0·99	0·87	1·061
λ_1	3·81	3·82	3·75	3·800
λ_2	7·36	7·52	7·41	—
λ_3	11·40	11·66	11·61	—
λ_4	—	16·27	16·23	—

The energy levels thus obtained can be checked by the Brillouin-Wentzel-Kramers (B.W.K.) method of approximation (Dunham 1932*a*). For any potential-energy function $V(x)$ the first two terms of the expansion are

$$(n+\tfrac{1}{2})h = \sqrt{(2m)}\oint\{W_n - V(x)\}^{\frac{1}{2}}\,dx - \frac{h^2}{64\pi\sqrt{(2m)}}\oint\left(\frac{dV}{dx}\right)^2\{W_n - V(x)\}^{-\frac{5}{2}}\,dx. \qquad (4)$$

Putting $V(x) = ax^4$ and $y = V(x)/W_n$, and using (2) it follows that

$$(2n+1)\pi = \tfrac{1}{4}\lambda_n^{\frac{3}{4}} I_1 - \tfrac{1}{8}\lambda_n^{-\frac{3}{4}} I_2, \tag{5}$$

$$\left.\begin{aligned} I_1 &= \oint (1-y)^{\frac{1}{2}} y^{-\frac{3}{4}}\,dy = 4\Gamma(\tfrac{3}{2})\,\Gamma(\tfrac{1}{4})/\Gamma(\tfrac{7}{4}) = 13{\cdot}98, \\ I_2 &= \oint (1-y)^{-\frac{5}{2}} y^{\frac{3}{4}}\,dy = 4\Gamma(\tfrac{7}{4})\,\Gamma(-\tfrac{3}{2})/\Gamma(\tfrac{1}{4}) = 2{\cdot}636. \end{aligned}\right\} \tag{6}$$

The values of λ_n calculated from equation (5) are given in column (*b*) of table 1. The agreement with column (*a*) is satisfactory except for the lowest level, which is just where the B.W.K. method is expected to be least reliable. A still simpler expression for the eigen-values of λ is obtained by omitting the last term of equations (4) and of (5), giving

$$\lambda_n = \left\{\frac{4\pi(2n+1)}{I_1}\right\}^{\frac{4}{3}}. \tag{7}$$

The values calculated from equation (7) are given in column (*c*) of table 1: except for the lowest level they are sufficiently accurate for most purposes. A further check on λ_0 and λ_1 is provided by the work of Milne (1930), who used a laborious but accurate numerical method, giving the values in column (*d*) of table 1. It may be noted that the addition of a third term to the B.W.K. expression (4) gives $\lambda_0 = 1{\cdot}15$, which differs more from the accurate value 1·061 than does the second approximation 0·99. This is consistent with the fact that the B.W.K. method involves an asymptotic and not a convergent expansion, and it is thus doubtful whether the inclusion of the second term in equation (4) produces in general an improvement on equation (7). The early divergence of the asymptotic series in the case $V(x) = ax^4$ can be related to the proximity of two imaginary zeros of the expression $W - V(x)$ (Kemble 1935). In subsequent calculations values in column (*d*) have been used for λ_0 and λ_1, those in column (*a*) for λ_2 and λ_3, and column (*c*) for λ_4 and higher levels.

It is interesting to note the change in the sequence of energy levels for the general symmetrical potential well $V(x) \propto |x^q|$ as the value of q is changed. Some of these values are given in table 2, the energy of the lowest level being taken as unity in each case.

TABLE 2

potential-energy function	energy levels				
$V(x) \propto \lvert x \rvert$	1	2·29	3·18	3·98	4·72
$V(x) \propto x^2$	1	3	5	7	9
$V(x) \propto x^4$	1	3·58	6·99	10·95	15·31
$V(x) = 0$ for $\lvert x \rvert < a$ $V(x) = \infty$ for $\lvert x \rvert > a$	1	4	9	16	25

The values in the first row have been derived elsewhere (Bell 1944), while the last row represents a rectangular potential box, corresponding to $q \to \infty$. These results may be compared with expressions previously derived for the permeability of a potential barrier with $V(x) = V_0(1 - a_q |x^q|)$, which show a similar gradation (Bell 1937). The asymptotic formula (15) shows that the gradation persists in the higher levels.

SELECTION RULES AND SPECTRUM

In the absence of explicit wave functions it is not possible to derive general selection rules for fourth-power vibrations. However, it is clear from symmetry considerations that the matrix element $\int_{-\infty}^{+\infty} \psi_m \psi_n x\,dx$ is zero when m and n are either both odd or both even, so that dipole transitions to or from the ground state ψ_0 can involve only odd upper states. Numerical calculation with the wave functions derived above gives $\int \psi_0 \psi_1 \xi\,d\xi = 1{\cdot}41$ and $\int \psi_0 \psi_3 \xi\,d\xi = 0{\cdot}015$. The latter value does not differ significantly from zero, which would suggest that the selection rule for dipole radiation is $n \to n \pm 1$, just as for the harmonic oscillator. However, it can be shown from the form of equation (3) that non-zero values of $\int \psi_m \psi_n \xi\,d\xi$ cannot always be restricted to $m = n \pm 1$,* and it seems likely that all transitions are allowed provided that m and n differ by an odd integer.

If the frequency for the transition $0 \leftrightarrow 1$ is taken as unity, then the frequencies of the overtones $0 \leftrightarrow 1$, $0 \leftrightarrow 2$, $0 \leftrightarrow 3$, $0 \leftrightarrow 4$, etc., are in the ratio 1, 2·32, 3·85, 5·54, etc., compared with 1, 2, 3, 4, etc., for a harmonic oscillator. Moreover, the transitions $1 \leftrightarrow 2$, $2 \leftrightarrow 3$, $3 \leftrightarrow 4$, $4 \leftrightarrow 5$, etc., will contribute other possible overtones with frequencies 1·32, 1·53, 1·69, 1·82, etc. The spectrum of a fourth-power oscillation is thus likely to contain more overtones than that of a harmonic vibration of comparable frequency.

In terms of the constant a the lowest spectrum frequency ν_0 is given by

$$\nu_0 = \tfrac{1}{4}(\lambda_1 - \lambda_0)\left\{\frac{ha}{\pi^4 m^2}\right\}^{\frac{1}{3}} = 0{\cdot}163\left(\frac{ha}{m^2}\right)^{\frac{1}{3}}. \tag{8}$$

It should be noted that this frequency is not identical (as for the harmonic oscillator) with the vibration frequency assigned by classical mechanics. This mechanical frequency ν' depends on the energy W, and is easily shown to be

$$\nu' = \frac{2\Gamma(\frac{3}{4})}{\sqrt{(2m\pi)}\,\Gamma(\frac{1}{4})}(Wa)^{\frac{1}{4}}. \tag{9}$$

For an energy W_0 corresponding to the lowest quantum state this becomes

$$\nu'_0 = \frac{\lambda_0^{\frac{1}{4}}\Gamma(\frac{3}{4})}{\pi^{\frac{5}{6}}\Gamma(\frac{1}{4})}\left(\frac{ha}{m^2}\right)^{\frac{1}{3}} = 0{\cdot}133\left(\frac{ha}{m^2}\right)^{\frac{1}{3}} = 0{\cdot}81\nu_0, \tag{10}$$

i.e. the 'fundamental' spectrum frequency is somewhat higher than the mechanical frequency of the system in its lowest state. According to the correspondence principle, as the quantum number n increases, the spectrum frequency for the transition $n \to n+1$ will approach more and more closely to the mechanical frequency of the system, and it is easily shown directly from equations (7) and (9) that this is so in the present case.

---

* The author is indebted to Professor E. C. Titchmarsh for pointing this out.

It is of interest to determine the approximate frequency of fourth-power vibrations in actual molecules, as a guide to their experimental detection. To do this the simple case of a square ring of four identical atoms is considered, assuming that forces between unlinked atoms can be neglected, and that the force constants for bond stretching are very large compared with those for valency bending. There are then two types of vibration which the ring can execute (without altering the bond lengths): (*a*) vibration in the plane of the ring, the atoms moving along the diagonals of the square, and (*b*) vibration at right angles to the plane of the ring, the atoms moving along the circles indicated in figure 1. For small displacements (*a*) is a simple harmonic vibration, while (*b*) is a fourth-power vibration. If Δ is the displacement of each atom, b the bond length, and d the ordinary bending force constant in dynes/cm., then the potential energy of the whole ring in the two cases is given by

$$V_a = 4\,.\,\tfrac{1}{2}\,.\,8d\Delta_a^2, \quad V_b = 4\,.\,\tfrac{1}{2}\,.\,\frac{16d\Delta_b^4}{b^2}. \tag{11}$$

Using the ordinary harmonic oscillator expression and equation (8), one finds for the lowest spectrum frequency in the two cases

$$\nu_a = \left\{\frac{2d}{\pi^2 m}\right\}^{\frac{1}{2}}, \quad \nu_b = \tfrac{1}{4}(\lambda_1 - \lambda_0)\left\{\frac{16hd}{\pi^4 m^2 b^2}\right\}^{\frac{1}{3}} = 1\cdot 37\left\{\frac{2hd}{\pi^4 m^2 b^2}\right\}^{\frac{1}{3}}, \tag{12}$$

and hence for the ratio of the two frequencies

$$\nu_b/\nu_a = 1\cdot 05\left\{\frac{h^2}{mdb^4}\right\}^{\frac{1}{6}}. \tag{13}$$

On putting reasonable values for m, d and b into equation (13), we find $\nu_b/\nu_a \simeq 0\cdot 7 - 0\cdot 2$. This value applies directly only to the simple model which is taken here, but it should be of the right order of magnitude in more complicated cases. Since the harmonic bending frequencies ν_a lie in the range 150–1500 cm.$^{-1}$ (according to the mass of the atoms), the fourth-power frequencies should be frequently detectable in the infra-red or Raman spectrum.

The molecule which approaches the model most closely is cyclobutane. If each CH_2 group can be considered to move as a whole, and interaction between unlinked atoms is neglected, then the second part of equation (12) should give an accurate value for the fourth-power frequency ν_b. Wilson (1943) has recently given an experimental and theoretical study of the vibrational spectrum of cyclobutane, obtaining a value of $3\cdot 28 \times 10^4$ dynes/cm. for the $C\hat{C}C$ bending force constant of the ring. With this value equation (12) gives $\nu_b = 120$ cm.$^{-1}$. This frequency is too low to be observed directly in the infra-red spectrum, but on the basis of observed combination tones Wilson has assigned the value 145 cm.$^{-1}$ to the out-of-plane bending frequency of the ring (which he regards as harmonic), which is somewhat larger than the figure calculated here. The discrepancy of 20 % suggests that restricted rotation is of appreciable but not decisive importance in treating this type of vibration. In the case of diborane the vibration spectrum has not yet been analysed on the basis of the bridge structure, but the value of ν_b can be estimated from equation (13). ν_a appears in the Raman spectrum as 793 cm.$^{-1}$, being identified by the boron isotope effect (Anderson & Burg 1938), while electron diffraction data give $b = 1{\cdot}39$ A (Longuet-Higgins & Bell 1943). Then from equation (13) it is found that $\nu_b \simeq 450$ cm.$^{-1}$, and there is in fact a strong infra-red band at 412 cm.$^{-1}$ which is otherwise difficult to explain (Stitt 1941). Close numerical agreement with equation (13) would not be expected, since it is certainly unjustifiable to neglect the interaction between unlinked boron atoms in calculating ν_a and ν_b.

Further identification of these frequencies as fourth-power vibrations might occur through their overtones, or through the thermal properties dealt with in the next section. However, the most direct verification would be by means of the isotope effect, since equation (8) shows that $\nu \propto m^{-\frac{2}{3}}$, instead of the usual $\nu \propto m^{-\frac{1}{2}}$. This test should be particularly valuable for diborane and similar compounds, where hydrogen atoms form part of the ring. In other cases the isotope effect can be approximately simulated by attaching groups to the ring, e.g. by comparing cyclobutane with alkyl-cyclobutanes. Since the fourth-power vibrations have a considerably lower frequency than the other vibrations of the system, the primary effect of substitution is merely to increase the mass of the units composing the ring.

In dealing with the spectrum and with the thermal properties, the fourth-power vibration is being treated as if it were one of the normal vibrations of the system, and thus qualified to make an independent contribution to the kinetic and potential energy of the system. This is not strictly legitimate, since the whole theory of normal vibrations depends on the assumption of harmonic vibrations of small amplitude. However, on account of the low frequency of the fourth-power vibration, it seems physically reasonable to treat it independently of the other modes of vibration, at least as a first approximation.

THERMODYNAMIC PROPERTIES

The contribution made by a fourth-power vibration to the various thermodynamic functions can be evaluated by the standard methods. For high temperatures equation (7) can be used in conjunction with the Euler-Maclaurin summation formula, but for lower temperatures it is necessary to use the more exact energy levels and to sum term by term. For a given fundamental spectrum frequency ν_0, the contribution of a fourth-power vibration to any thermodynamic function will differ from that of a harmonic vibration. The extent of the deviation is illustrated by table 3, which gives the molar specific heat as a function of $h\nu_0/kT$. It will be

TABLE 3. SPECIFIC HEAT CONTRIBUTIONS

(a) = harmonic vibrations; (b) = fourth-power vibrations

$h\nu_0/kT$		0	1·0	1·5	2·0	3·0	4·5	6·0
c_v/R	(a)	1·000	0·921	0·832	0·724	0·496	0·230	0·0898
	(b)	0·750	0·802	0·685	0·620	0·460	0·219	0·0897

seen that at low temperatures there is little difference between the two types of oscillator, corresponding to the fact that only the two lowest levels are effective in determining the specific heat. At higher temperatures the difference is considerable, and should be detectable experimentally. At intermediate temperatures the value of c_v/R for the fourth-power vibration rises slightly above the limiting high-temperature value of 3/4. This limiting value is in accordance with the classical equipartition theory, according to which any form of energy proportional to the qth power of a co-ordinate or a momentum contributes RT/q to the total energy per mole: in the present instance the kinetic energy contributes $\frac{1}{2}RT$ and the potential energy $\frac{1}{4}RT$.

It is interesting in this connexion to consider the more general oscillator represented by

$$V(x) = a_q\,|\,x^q\,|, \tag{14}$$

where q can have any positive value. For high quantum numbers the energies are given by the first term of equation (4), which reduces to

$$W_n^{1/q+\frac{1}{2}} = \frac{q(n+\frac{1}{2})\,ha_q^{1/q}}{2\sqrt{(2m\pi)}}\,\frac{\Gamma(1/q+\frac{3}{2})}{\Gamma(1/q)}. \tag{15}$$

At high temperatures n can be treated as a continuous variable and the summations in the partition function replaced by integrations. Putting $t = W_n/kT$ this gives for the mean energy per mole

$$\bar{E} = \frac{RT\displaystyle\int_0^\infty t^{1/q+\frac{1}{2}}\,e^{-t}\,dt}{\displaystyle\int_0^\infty t^{1/q-\frac{1}{2}}\,e^{-t}\,dt} = \left(\frac{1}{q}+\frac{1}{2}\right)RT, \tag{16}$$

in agreement with the classical result.

REFERENCES

Anderson, T. F. & Burg, A. B. 1938 *J. Chem. Phys.* **6**, 586.
Anderson, T. F., Lassettre, E. N. & Yost, D. M. 1936 *J. Chem. Phys.* **4**, 703.
Aston, J. G., Fink, H. L. & Schumann, S. C. 1943 *J. Amer. Chem. Soc.* **65**, 341.
Bauer, S. H. & Beach, J. Y. 1942 *J. Amer. Chem. Soc.* **64**, 1147.
Bell, R. P. 1937 *Proc. Roy. Soc.* A, **158**, 128.
Bell, R. P. 1944 *Phil. Mag.* (to be published shortly).
Brode, H. 1940 *Ann. Phys., Lpz.* (5), **37**, 344.
Dunham, J. L. 1932*a* *Phys. Rev.* **41**, 713.
Dunham, J. L. 1932*b* *Phys. Rev.* **41**, 721.
Kemble, E. C. 1935 *Phys. Rev.* **48**, 549.
Kimball, G. E. & Shortley, G. H. 1934 *Phys. Rev.* **45**, 815.
Longuet-Higgins, H. C. & Bell, R. P. 1943 *J. Chem. Soc.* 250.
Mann, F. G. 1938 *Ann. Rep. Chem. Soc.* **35**, 148.
Milne, W. E. 1930 *Phys. Rev.* **35**, 863.
Palmer, K. J. & Elliott, N. 1938 *J. Amer. Chem. Soc.* **60**, 1852.
Stitt, F. 1941 *J. Chem. Phys.* **9**, 780.
Wilson, T. P. 1943 *J. Chem. Phys.* **11**, 369.

Absorption of radium (B + C) gamma-rays

By J. E. Roberts, D.Sc., F.Inst.P.
Barnato Joel Laboratories, The Middlesex Hospital

(*Communicated by E. C. Stoner, F.R.S.—Received* 12 *January* 1944)

New measurements of the absorption of filtered gamma-rays from radium (B + C) in aluminium, carbon and lead have been made. A small condenser type of ionization chamber has been used, which overcomes many of the difficulties usually inherent in this kind of measurement. Detailed consideration has been given to the corrections which must be applied to ionization measurements before absorption coefficients can be calculated.

From the most recent theories of absorption by scattering, photoelectric effect, and electron pair formation, total absorption coefficients for lead over the wave-length range concerned have been calculated. The problem of the effective wave-length of a filtered heterogeneous gamma-ray beam is discussed in some detail, and a proposed method of estimation put forward. Comparison of the calculations of μ/ρ with the experimental figures indicates that the photoelectric absorption rises more rapidly with increasing wave-length than is predicted by theory.

INTRODUCTION

A number of attempts have been made to determine experimentally the absorption coefficients in lead and other materials of the gamma-rays of radium (B + C), under varying degrees of filtration. Although experimental methods do not differ fundamentally, the results obtained have not been in very good agreement. The classical work on the subject was carried out by Kohlrausch (1917), and his results have been quoted and used by most workers. Since that time Ahmad (1924*a*, 1924*b*, 1925) has made a detailed investigation of the absorption, in a number of materials, of gamma-rays of a fixed quality (1 cm. lead filter). The results of these and other experimenters have been summarized by Rutherford, Chadwick & Ellis (1930).

The calculation of absorption coefficients from the experimental data involves the evaluation of corrections connected with scattering and with energy absorption in the ionization chamber. With the usual experimental arrangements these corrections may be large and are difficult to estimate with precision. This leads to some uncertainty in the absolute values of the absorption coefficients. In recent years, due to a large extent to the development of the medical uses of radium, considerable advances have been made in methods of measurement of gamma-rays, and in knowledge of the absorption processes. The purpose of the present work has been to use these methods in an attempt to eliminate the disturbing factors as far as possible, and where this could not be done, to make reasonably accurate allowances for them. The experiments have been limited to the gamma-rays of radium (B + C), with filtrations up to 5 cm. of lead, using absorbers of lead, aluminium and carbon. Having obtained the absorption coefficients, a further attempt is made to analyse the results and correlate them with the most recent theories of absorption.

APPARATUS

General arrangement

The general arrangement of the apparatus for absorption measurements follows the principles laid down by Kohlrausch (1917). A beam of gamma-rays from a radium source R (figure 1) is canalized by means of a series of heavy lead stops before falling on the ionization chamber C. With the ionization system described below, a high degree of canalization was not required. There is no electrometer or other accessory apparatus in the vicinity, and it was only necessary to prevent excessive scattered radiation from surrounding objects from reaching the ionization chamber. The stops were built up of lead 'bricks', S_1 being 8 cm. thick, and S_2 and S_3 each 4 cm. thick. The aperture in S_1 was 4 cm. square, and in the other stops 3 cm. square. Filters up to 5 cm. thick could be placed between S_1 and S_2, while S_2 and S_3 were separated by 2·5 cm. to take the absorbers. Between S_3 and the ionization chamber a heavy lead shutter could be interposed. This consisted of a cube of 10 cm. side mounted on a smooth running trolley, and could be moved quickly in and out of position by hand. The whole system was mounted on a wooden table in the middle of a large room, thus further reducing the possible effects of stray scattered radiation.

A single source with small dimensions and of sufficient strength not being available, a composite radium source was used. This was made up of ten platinum tubes of the type used in medical work, and had a total radium content of approximately 100 mg. The tubes were mounted on a small brass plate, the radium being spread over an area of about 3 cm. square. The '$\cos\theta$ error' introduced into the measurements by this dispersal was negligible. The wall thickness of the radium tubes was 1 mm., and this set one limit to the qualities of radiation available. For most purposes this 1 mm. of platinum was regarded as equivalent in absorption to 2 mm. of lead.

In order to vary the quality of the gamma-ray beam, lead filters were placed in position F (figure 1). A series of eight such filters was used, with thicknesses up to 4 cm., and made up for the most part of sheets of lead 5 cm. square and 0·5 cm. thick. Greater filtrations were possible, but the change in radiation quality obtained hardly justified the additional work involved.

The absorbing materials to be investigated, in the form of 5 cm. square blocks, were placed in position A (figure 1). The main characteristics of the absorbers are shown in table 1.* A density of 11·36 g./cm.^{3} indicates a high degree of purity in the lead; the aluminium was high-quality commercial material; the carbon absorber was a block of Acheson graphite.

TABLE 1

absorber	thickness cm.	density g./cm.^{3}	superficial mass g./cm.^{2}
lead I	0·500	11·36	5·67
lead II	0·167	11·36	1·90
aluminium	1·931	2·71	5·23
carbon	2·00	1·58	3·16

---

* I am indebted to Mr B. D. H. Watters for accurate measurements of the absorbers.

Ionization system

The choice of an ionization system for precision measurements of gamma-rays at a large distance from the radium source presents considerable difficulties, largely owing to the very small currents involved. With the arrangement shown in figure 1, and using a 100 mg. source, the ionization per c.c. of chamber volume amounts to only 4×10^5 ions/sec. without filtration, falling to 2×10^4/sec. with 4 cm. lead filter. This gives currents between 6×10^{-14} and 3×10^{-15} amp./c.c. of ionized air. To overcome these difficulties, it has been the usual practice to use a very sensitive measuring instrument, together with an ionization chamber of large volume. The latter inevitably introduces difficulties in the interpretation of the results. Some method of integrating the ionic current over a period is clearly desirable. Developments in recent years of the measurement of gamma-ray dosage for medical purposes have shown the very great advantages of the condenser type of ionization chamber introduced by Glasser (1928) and Sievert (1932). This consists essentially of a small cylindrical or spherical condenser, with a hole in the outer electrode for charging and measuring purposes. The inner electrode is charged to a known potential, the chamber closed, and then placed in position for irradiation. After irradiation

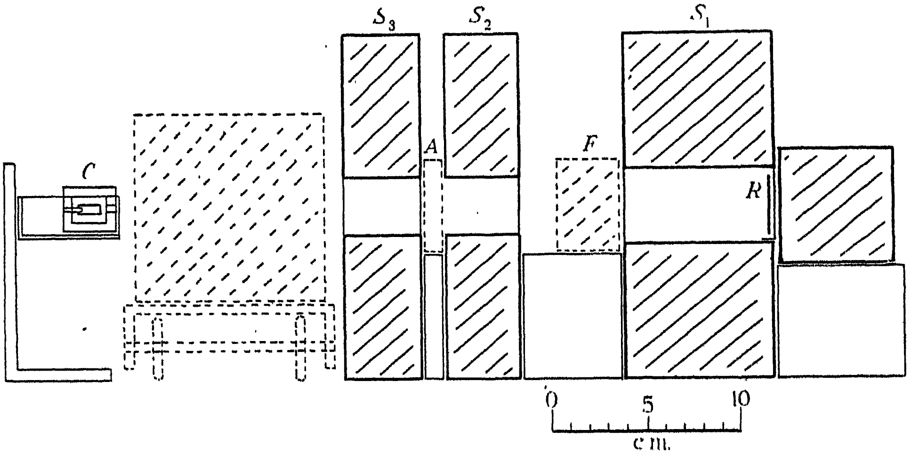


FIGURE 1. Lay out of apparatus.

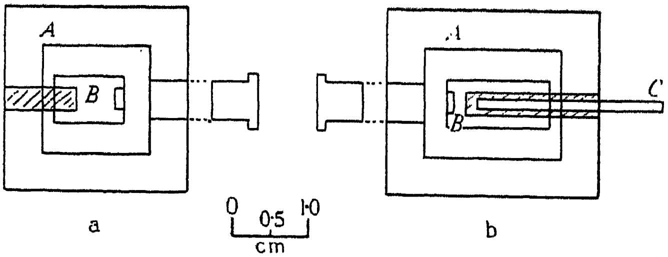


FIGURE 2. Ionization chambers.

for a known time the residual potential on the inner electrode is measured by sharing the charge with a calibrated electrometer. So long as the residual potential does not fall below the saturation value for the chamber, the difference between the initial and final potentials is a measure of the ionization produced in the air volume. No measurements are taken during irradiation. There are no leads to be protected from stray ionization, and the electrometer can be kept well away from the radium, again requiring no protection. A further advantage is that if a wide range of intensities has to be investigated, the magnitudes of the measured quantities can be kept roughly constant by varying the exposure time. A high degree of insulation between the electrodes is of course necessary, as a guard ring cannot be used.

For the experiments here described, two types of condenser chamber were used, and they are shown in figure 2. The simplest form (figure 2*a*) is as described above. The outer electrode *A* consists of a cylinder of Acheson graphite with internal dimensions 1·3 cm. diameter and 1·3 cm. long, and wall thickness 0·5 cm. It is necessary to use a wall of this thickness (the maximum range of the photo-electrons produced by the radiation) in order that the ionization measured should be proportional to the incident gamma-ray energy (Mayneord & Roberts 1937). A 5 mm. hole in the end of the cylinder for charging purposes is closed by a graphite plug. The inner electrode, *B*, is suspended concentrically with the outer by means of a small Distrene (thermoplastic polystyrene) peg. It also is made of graphite, 0·9 cm. long and 0·6 cm. diameter, and has a small button of Elektron metal at the lower end to prevent damage to the graphite during charging. The procedure of measurement was as follows. With the outer electrode earthed, the central electrode is brought into contact with an insulated steel needle, raised to about 120 V. This initial potential is measured on a precision voltmeter. The chamber is then closed and placed in a V-shaped holder in a fixed position in the gamma-ray beam. The time of exposure can be adjusted so that whatever the radiation intensity, the final voltage of the electrode is in the region of 30 V. In these experiments the exposures ranged from 10 min. to 4 hr., the timing errors being extremely small. After irradiation the chamber is removed, and the inner electrode again brought into contact with the steel needle which is now connected to a Lindemann electrometer. The latter is calibrated at a fixed sensitivity, using the same voltmeter as for the initial potential. The difference between the initial and final electrode potentials is a measure of the total ionic charge collected during the exposure. For most experiments an electrometer sensitivity of 30 div./V was sufficient, the capacity of the electrometer system being about twelve times that of the chamber.

In the second type of chamber (figure 2*b*) a third electrode, *C*, was introduced. The insulating peg now extends deep into electrode *B*, and is drilled to take a 1 mm. thick wire which projects from the back of the chamber. With this arrangement all the electrodes are initially earthed, *B*, by contact with the steel needle. On being placed in the V holder, electrode *C* is kept at 120 V with respect to *A*. The ratio of the capacities *A* to *B* and *B* to *C* is such that there is a potential difference of some 80 V across the air space between *A* and *B*. On irradiation, electrode *B* collects the ionic charge, and this is then shared with the electrometer. At the time of sharing, a compensating voltage, finely adjusted by means of a potentiometer, is applied between *C* and *A*. A null reading is therefore obtained. The compensating voltage, read on a precision voltmeter, gives a direct measure of the integrated ionization during the exposure. Using this chamber, much shorter exposures were possible, but, for the sake of accuracy in timing, were kept in the region of 10 min. No significant differences were found between the results from the two arrangements, but somewhat greater reliance has been placed on those obtained with the second type of chamber.

A large number of measurements were made to test the consistency of the ionization readings obtained with these chambers. For any one set of conditions, the maximum deviation from the mean value rarely exceeded 4 parts in 1000, and the final mean values for the ionizations are probably correct to 2 parts in 1000.

Owing to the very low radiation intensities, maintenance of saturation conditions in the chambers did not present any difficulties. To test the saturation, the chambers were exposed for increasing periods to the maximum radiation intensity used. With chamber *a* (figure 2*a*), the measured intensity (in volts loss per minute) remained sensibly constant until the residual potential fell to 16 V. In practice it was not allowed to fall below 30 V. Chamber *b* (figure 2*b*) will become unsaturated if the charge collected on electrode *B* reduces the potential difference between *A* and *B* too far. The experiments showed no indication of this happening over the range considered. At lower radiation intensities saturation will be even more complete.

As some of the series of measurements were spread over a fairly long period of time, it was necessary to take into account variations in the density of the air in the ionization chamber. All ionization readings were therefore reduced to a standard temperature and pressure. An allowance was also made for the slight loss of charge from the chamber due to insulation leak and natural ionization. To estimate the leak, the chamber was charged to a point midway between the usual initial and final potentials, and allowed to stand for periods up to 24 hr. in the absence of radium. The loss of potential per hour was then observed, and the appropriate correction applied to the ionization readings. For chamber *a* the usual leak was about 0·3V/hr., so that even on the longest exposures the correction was little more than 1 %. Owing to the shorter exposures used with chamber *b*, the natural leak corrections never exceeded 0·2 %, and were usually much less.

DETERMINATION OF ABSORPTION COEFFICIENT

The ideal method for the determination of an absorption coefficient is to measure the slope at the appropriate point on a logarithmic transmission curve in the material considered. In practice, this method presents a number of difficulties, particularly with a heavy element such as lead. The quality of the radiation reaching the ionization chamber varies considerably with the filtration, and to obtain a true energy transmission curve, it is necessary to know precisely the constitution of the beam at all stages, and the ionization function of the chamber. Exact correction for the amount of scattered radiation reaching the chamber is also difficult if a variable absorber thickness is used. Furthermore, there is the purely geometrical difficulty of finding the slope of a curve which in some parts is varying only very slowly. In view of these objections, the method of Ahmad and others, of measuring the percentage transmission through a fixed absorber, of any given radiation beam, has been adopted, the coefficient then being determined by direct calculation. In the case of lead, it was necessary to use two absorbers. With lightly filtered radiation, passage through the 5 mm. lead absorber changed the quality of the beam appreciably, and the measured coefficient was found to decrease somewhat with increasing absorber thickness. For this reason, in the measurements on more heterogeneous beams, the 5 mm. lead absorber was replaced by one of 1·67 mm., giving a closer approximation to the true absorption coefficient.

The choice of absorber thickness is also very closely bound up with the accuracy to which the coefficients can be determined. If x is the probable percentage error in one ionization measurement, then in the percentage transmission there is a probable error of $\sqrt{(2)}\,x$. Assuming exponential absorption, to a first approximation it is clear that the error in the absorption coefficient is $\sqrt{(2)}\,x/F$, where F is the fraction of radiation absorbed. The lead and aluminium absorbers in the present experiments were chosen to transmit 70–75 % of the radiation. Thus, assuming the ionization measurements are correct to 0·2 %, the values of the absorption coefficients should have a probable error of not more than $1\frac{1}{2}$ %. In the case of graphite the transmission was 80–85 %, giving an error of some $2\frac{1}{2}$ %. It can also be seen that the use of the thin lead absorber, except for very heterogeneous beams, although desirable from some points of view could not be justified, owing to the loss in accuracy. It may be pointed out here that the accuracy of the ionization measurements does not depend to any extent on the radiation intensity. Approximately the same ionic charge is measured each time. This is not the case with the more usual direct measurement of ionization current.

ABSORPTION COEFFICIENTS IN CARBON, ALUMINIUM AND LEAD

Apparent absorption coefficients and corrections

The apparent mass absorption coefficients (μ/ρ) for carbon and aluminium, as calculated directly from the measured ionization currents, with and without the absorber in position, are shown in the broken curves of figure 3*b*. In the case of lead, the apparent coefficients, thus calculated, are given in column 2 of table 2. To convert these values into energy absorption coefficients in the strict sense, two important corrections must be applied. These involve estimations of the scattered radiation reaching the chamber from filter and absorber, and of the ionization function of the chamber.

Scattered radiation

The forward scattering of radiation into the ionization chamber by the absorber may lead to appreciable errors in the determination of the transmission through the absorber. Kohlrausch partially overcomes this by using large source-absorber-chamber distances, but this leads to very low radiation intensities. On the other hand, Ahmad, by taking readings at two absorber positions subtending 20° and 80° at the ionization chamber, extrapolates the results to a condition of zero scattering. Although there are unavoidable uncertainties in this calculation, the results agree fairly well with those of Kohlrausch.

Tarrant (1932) has shown that, with the arrangement of figure 1, the correction for scattering is given by

$${}_e\mu_{\text{true}} - {}_e\mu_{\text{exp.}} = \frac{\pi e^4}{m^2c^4}\,\theta^2\{1+\tfrac{1}{4}\theta^2(\tfrac{11}{2}\alpha-1)\},$$

where $\alpha = h\nu/mc^2$ and θ is the maximum angle of scatter from absorber to chamber. In order to make an approximate check of this formula, two series of absorption measurements were made using different absorber-chamber distances to give $\theta = 9\tfrac{1}{2}°$ and $\theta = 14\tfrac{1}{2}°$. When corrected in this way for scatter, the values for μ/ρ for the two distances were found to agree within the experimental limits. For the purposes of this correction, the approximate effective wave-lengths of the radiations were estimated from the aluminium absorption data of Kohlrausch, using the Klein-Nishina formula to connect absorption coefficient and wave-length. Most of the main experiments were carried out with the absorber subtending a half angle of $9\tfrac{1}{2}°$, though a few results with $\theta = 14\tfrac{1}{2}°$ are included.

The Tarrant correction for scattering takes no account of the absorber thickness, which may reduce the amount of its own scattered radiation. For the 5 mm. lead and the aluminium absorbers, therefore, a further step in the correction is necessary. Tandberg (1937) has shown that the true transmission through a thick absorber is given by

$$\frac{I_d}{I_0} = e^{-\mu d}\,e^{-\beta},$$

where the correction factor $\beta = (k_\theta/\mu)(1-e^{-\mu d})$ and k_θ is the fraction of energy scattered between 0° and $\theta°$, the angle subtended by the absorber at the chamber. This formula reduces to that of Tarrant at small absorber thicknesses. According to Tarrant's calculation from the Klein-Nishina formula, the value of k_θ per electron is given by

$$\frac{\pi e^4}{m^2c^4}\,\theta^2\{1+\tfrac{1}{4}\theta^2(\tfrac{11}{2}\alpha-1)\}.$$

To assess the value of β, it is necessary to assume approximate values for μ, and the experimental figures, as corrected by the Tarrant formula, were used. As is to be expected, the correction from the Tandberg formula is somewhat smaller than from that of Tarrant.

Some scattered radiation from the filter will also reach the ionization chamber. It will, however, form less than 2 % of the beam incident on the absorber, and will have, on the average, a wave-length longer than the main beam by about 0·015 A. The effect of this radiation on the quality of the primary beam can clearly be neglected.

Ionization function of chamber

The estimation of the ionization function of the chamber is perhaps the most troublesome of all the aspects of precise gamma-ray measurements. This involves the determination of the relationship between measured ionization and the energy incident on the chamber. This relationship depends on the absorption in the chamber wall and in the enclosed air, which, in turn, depend on the wall material and the wave-length of the incident radiation. L. H. Gray (1936) has investigated this problem and shown that, in a small air cavity situated in a block of solid material, the rate of ionization I in the cavity is related to the energy flux E by

$$I = \frac{E({}_e\sigma_a + {}_e\tau)\,n}{W\rho_e},$$

where ${}_e\sigma_a$ and ${}_e\tau$ are electronic scattering and photoelectric absorption coefficients for the wall material,

n is the number of electrons per unit volume of air,

W is the energy required to produce one ion pair,

ρ_e is the ratio of the electronic stopping powers of the wall material and the gas.

This relationship is subject to the conditions that the dimensions of the air cavity are small compared to the photoelectron range in air, and that the wall thickness is at least equal to the maximum range of electrons produced in the wall material. It is clear that the large type of electroscope or large-volume ionization chamber commonly used for gamma-ray measurements cannot fulfil these conditions. In an analysis of Kohlrausch's experimental results, Sizoo & Williams (1938) have assumed that all the ionization was due to air absorption. This is unlikely to be correct, as the design of the chamber is such that the lead walls must contribute appreciably to the ionization. Ahmad (1924*a*) attempted to absorb the greater part of the radiation in the ionization chamber by introducing iron and lead plates. As has been pointed out by Stoner (1929), this does not entirely answer the problem, and the calculation of the ionization function is greatly complicated. Stoner (1929) has carried out this calculation with reasonably satisfactory results. On the other hand, J. A. Gray (1927), using a wood and graphite electroscope, measured the absorption of a graphite plate in contact with the chamber wall, and then used the coefficient obtained as the ionization function of the electroscope. This approximates more nearly to the method suggested by the theory of L. H. Gray.

The use of a small condenser type of ionization chamber makes it possible to fulfil quite closely the requirements of the above equation. The 5 mm. wall thickness is greater than the maximum range of the secondary electrons produced in graphite, and the linear dimensions of the air volume are of the order of 1 % of the average electron range in air.

L. H. Gray has shown that W and ρ_e are sensibly constant over a wide range of wave-lengths. Furthermore, for a graphite chamber and the gamma-rays of radium (B + C), photoelectric absorption in the wall is negligible. One can therefore write

$$\frac{I}{E} = \text{const.} \times {}_e\sigma_a.$$

Using the experimental method for the determination of μ/ρ described earlier, this ratio I/E only varies when the mean wave-length of the radiation is changed during absorption. In the cases of carbon and aluminium absorbers this change is negligible, and no correction of the values of μ/ρ for ionization function is necessary. For the lead absorbers we obtain the percentage energy transmission

$$\frac{E}{E_0} = \frac{I}{I_0}\frac{{}_e^0\sigma_a}{{}_e\sigma_a},$$

where ${}_e^0\sigma_a$ corresponds to the radiation before absorption, and ${}_e\sigma_a$ corresponds to the radiation after absorption. The effective wave-lengths for the determination of ${}_e\sigma_a$ were obtained by the method described later. Although ${}_e\sigma_a$ varies only relatively slowly with wave-length, the effect of this variation is greatly magnified in the calculation of absorption coefficients. For the most lightly filtered radiations, the change from ionization to energy ratios may lead to a change of as much as 10 % in the value of μ/ρ for lead. The magnitudes of the scattering and ionization function corrections at a number of filtrations are indicated in table 2. The correction for scattering in the cases of carbon and aluminium absorbers can be seen from figure 3 *b*.

TABLE 2. CORRECTIONS TO ABSORPTION COEFFICIENTS IN LEAD

filter cm. lead	μ/ρ (apparent)	μ/ρ corrected for scatter only	μ/ρ corrected for scatter and ionization function
0·2	0·0985	0·1002	0·0909
0·7	0·0680	0·0697	0·0658
1·2	0·0601	0·0650	0·0584
1·7	0·0559	0·0572	0·0550
2·2	0·0527	0·0541	0·0519
3·2	0·0479	0·0511	0·0498
4·2	0·0470	0·0485	0·0474

Corrected absorption coefficients

The finally corrected values for the total mass absorption coefficients (μ/ρ) for aluminium, carbon and lead are plotted as a function of lead filter thickness in figure 3 *a*, *b*. In view of the rather different approach to the problem, the results for aluminium agree well with those calculated by Rutherford from the data of Kohlrausch. The new values of μ/ρ for lead are appreciably lower than those given by

Rutherford over the whole range. This is very probably due to a more adequate compensation in the present results for chamber ionization function. Lack of this compensation will, in general, tend to give high values for μ/ρ.

Although the values of μ/ρ for carbon are subject to greater experimental error than the rest, a comparison of the results for aluminium and carbon forms a useful

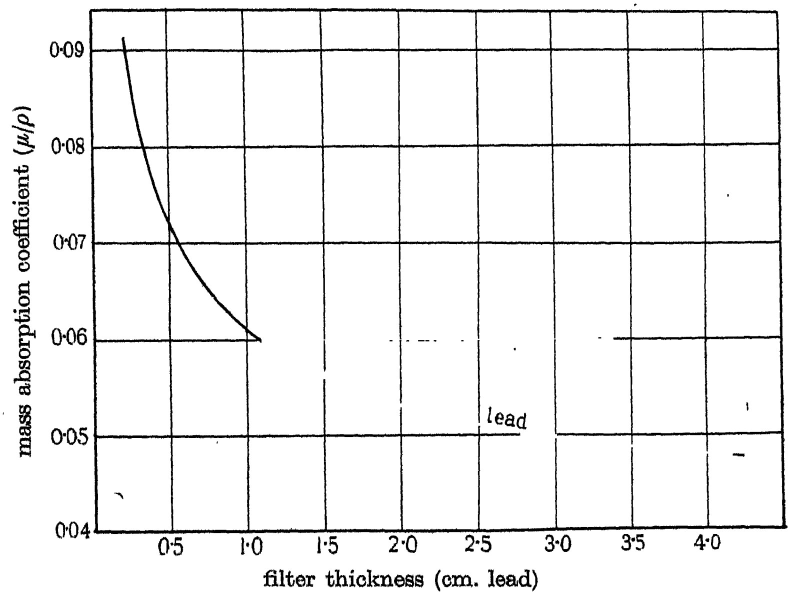


FIGURE 3*a*. Absorption of gamma-rays in lead (corrected values).

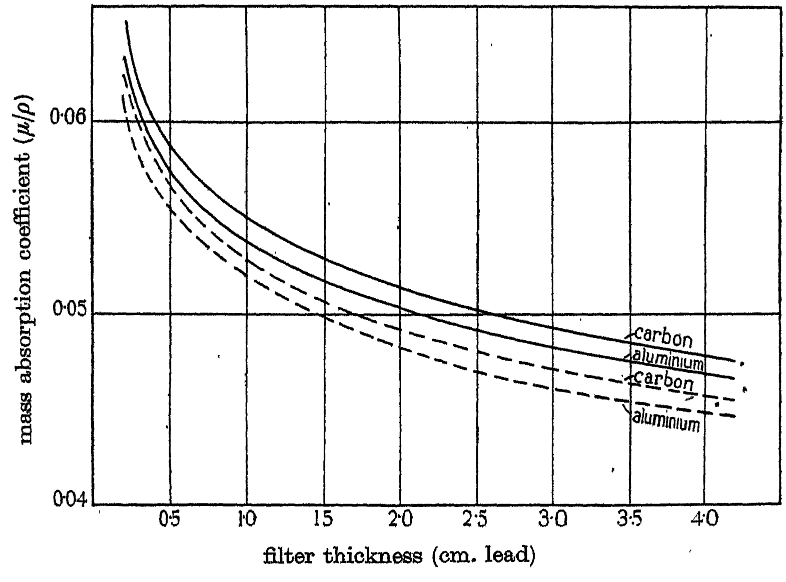


FIGURE 3*b*. Absorption of gamma-rays in carbon and aluminium. Broken curves, apparent values; full curves, corrected values.

check on the method. In table 3 are shown the values of the absorption per electron ($_e\mu$) for a number of radiation qualities. The absorption in carbon and aluminium in this wave-length range is almost entirely by scatter, and hence the absorption per electron should be the same for both elements. Table 3 shows that this is the case to well within the experimental error. The small difference between the μ/ρ curves for carbon and aluminium is thus due to the difference in the electronic densities of the two materials.

TABLE 3. ELECTRONIC ABSORPTION COEFFICIENTS IN CARBON AND ALUMINIUM

filter, cm. lead	0·2	0·5	1·0	2·0	3·0	4·0
$_e\mu \times 10^{24}$ carbon	0·216	0·195	0·183	0·171	0·164	0·159
$_e\mu \times 10^{24}$ aluminium	0·217	0·197	0·184	0·172	0·165	0·161

ABSORPTION COEFFICIENT AND WAVE-LENGTH

Structure of the absorption coefficient

As there is no photonuclear absorption in the materials used, the total absorption coefficient may be written as

$$\frac{\mu}{\rho}=\frac{\sigma}{\rho}+\frac{\tau}{\rho}+\frac{\chi}{\rho},$$

where σ/ρ is the coefficient corresponding to scattering,
τ/ρ corresponds to photoelectric absorption,
χ/ρ corresponds to electron pair formation.

Each of these factors has been investigated theoretically by various authors. Scattering of high-frequency radiation is well accounted for by the formula of Klein & Nishina (1928), from which values of σ can be calculated.* The validity of the formula over a wide range of wave-lengths has been established by Read & Lauritsen (1936) and others using monochromatic X-rays and gamma-rays. The most recent theoretical treatment of the problem of photoelectric absorption is that of Hulme, McDougall, Buckingham & Fowler (1935). They have considered the perturbation of a simple atomic system by a gamma-ray beam, and, with some approximations, have obtained curves connecting the photoelectric coefficient τ with the wave-length of the incident beam. Their results are given as graphs of $\tau_{\text{atom}}/Z^5\theta$ against θ, where $\theta = mc^2/h\nu$. For the present purposes the Hulme curve for lead has been used to calculate the values of τ/ρ up to a wave-length of 35 X.U., and extrapolated with the aid of L. H. Gray's (1931) empirical formula up to 50 X.U. Absorption by electron pair formation is only significant for the shorter components of the gamma-rays of radium (B + C), and forms a small percentage of their total absorption. Approximate values of χ in this region have been obtained from the theoretical work of Bethe & Heitler (1934). The results of calculations for lead from these three sources are shown for wave-lengths up to 16 X.U. in figure 5*a*. Further figures over a wider wave-length range are given in column 2 of table 5.

---

* The writer is indebted to Mr S. B. Osborn, who has recalculated the values of $_e\sigma$ to a high degree of accuracy, using the most recent values of the atomic constants.

Effective wave-lengths

Before the experimental results given in figure 3 can be compared with theoretical predictions, it is necessary to make some estimate of the effective wave-lengths of the gamma-ray beams used. Except under extremely heavy filtrations the gamma-rays of radium (B + C) are very heterogeneous, and it does not appear possible to determine a single effective wave-length suitable for all purposes. The most obvious method, as used by a number of workers, is to make use of the aluminium absorption coefficients of the same beams as are used for the measurements on lead. Absorption in aluminium in this region being purely by scatter, the scattering per electron (${}_e\sigma$) can be derived for each radiation quality from the experimental figures. The corresponding wave-lengths are then found from the Klein-Nishina formula. These wave-lengths have, however, a limited application. Strictly they should only be used when absorption by scatter alone of the heterogeneous beams is being investigated. Photoelectric absorption increases much more rapidly with wave-length than scatter, and hence with a heterogeneous beam the longer wave-lengths predominate in determining the coefficients. This leads to a higher effective wave-length. In the case of radium gamma-rays in lead, the problem is more complicated, as both scattering and photoelectric absorption play a part.

In view of these difficulties, an attempt has been made to adapt the method indicated by Stoner (1929), and to calculate the effective wave-lengths of filtered gamma-rays from the known data on the spectrum of radium (B + C). This spectral distribution has been investigated by different methods by Skobelzyn (1927, 1929) and Ellis & Aston (1930).

Using an expansion chamber, Skobelzyn measured the numbers and velocities of the recoil electron tracks produced within 20° of a narrow pencil of gamma-rays. From each such track the frequency of the original quantum producing it could be deduced. Thus as, according to the Klein-Nishina formula, the number of electrons ejected within 20° of the gamma-ray beam is practically independent of the frequency, the number of tracks corresponding to any one wave-length gives a measure of the relative number of quanta of that wave-length in the beam. When Skobelzyn's results are plotted as a distribution curve against wave-length, a series of well-defined peaks are observed. Most of these can be quite definitely associated with the known β-ray spectrum of radium (B + C). It is then a comparatively simple matter to separate the peaks and so estimate the relative number of quanta corresponding to each wave-length in the gamma-ray spectrum. These are shown in column 3 of table 4. To convert these figures into the energy distribution it is only necessary to multiply each value of ΔN by the appropriate $h\nu$. In column 5 of table 4, the energy ΔE of each spectrum line is given as a percentage of the total emission.

Ellis & Aston (1930) determined, by a photographic method, the relative photo-electron emission from platinum for each observable line in the gamma-ray spectrum. Their results are given in terms of $N\tau_k$ for each wave-length, where N is the relative number of quanta in that wave-length and τ_k is the corresponding k level photo-electric absorption coefficient in platinum. To obtain N it is therefore necessary to make some assumption as to the variation of τ with wave-length, but as a first approximation the formula of Hulme, McDougall, Buckingham & Fowler (1935) may be used. For the present purpose this is a disadvantage, making the final comparison of experimental and theoretical results rather uncertain. In this attempt to determine effective wave-lengths, both spectral distributions have been investigated, but to avoid confusion, only the Skobelzyn figures are shown in the accompanying tables.

TABLE 4. ENERGY DISTRIBUTION IN GAMMA-RAY SPECTRUM (FILTRATION, 0·4 CM. LEAD)

λ X.U.	$h\nu \times 10^3$ V	number of tracks ΔN	$h\nu \times \Delta N = \Delta E$	ΔE %
4·15	2980	2	6·0	1·2
4·64	2660	6	16·0	3·1
5·10	2420	6	14·5	2·7
5·57	2219	14	31·1	6·1
5·91	2085	13	27·1	5·3
6·94	1778	69	122·7	24·0
8·89	1390	65	90·4	17·7
9·89	1248	24	30·0	5·9
10·92	1130	29	32·8	6·4
13·1	941	35	32·9	6·5
16·0	773	26	20·1	3·9
20·2	612	105	64·3	12·6
34·9	354	27	9·6	1·9
41·6	297	29	8·6	1·7
50·8	243	21	5·1	1·0

In order to obtain the energy distribution for each filter used in the experiments, one has to reduce ΔE for each wave-length by the absorption factor ($e^{-\mu d}$) appropriate to the filter. The coefficients (μ) calculated from theory in the previous section have been used. The results of the calculations, giving energy distributions up to 5 cm. lead filtration are shown in table 5. Incidentally, it may be pointed out that for each filtration, the sum $\Sigma\Delta E$ gives the percentage transmission of the composite radiation, and a predicted transmission curve can thus be built up. The filtration used by Skobelzyn (3·5 mm. lead + 2 mm. glass) is taken as equivalent to 4 mm. lead.

Using the energy distributions shown in table 5, it is now possible to estimate the effective wave-length of the radiation for each filtration. The simplest method, as suggested by Mayneord (1933), would be to weight each component according to the energy in that wave-length. Thus

$$\lambda_{\text{eff.}} = \frac{\Sigma\lambda\Delta E}{\Sigma\Delta E}.$$

This takes no account of the type of absorption to which the radiation is to be subjected. If the absorber is a light element, and photoelectric absorption is negligible, the effect of each energy component must be weighted accordingly. From the Klein-Nishina formula it can be shown that, over the wave-length range 5–50 X.U., the scattering absorption coefficient σ is very nearly proportional to $\lambda^{\frac{1}{2}}$. Thus, for such an absorber, it is probably more correct to write

$$\lambda_{\text{eff.}} = \left(\frac{\Sigma\lambda^{\frac{1}{2}}\Delta E}{\Sigma\Delta E}\right)^2.$$

Effective wave-lengths, calculated in this way, are shown in curve A of figure 4. It was found that the wave-lengths in this curve agreed reasonably well with those obtained from the experimental absorption data for aluminium and carbon, in which all the absorption is by scattering. The wave-lengths thus calculated were used in determining the chamber ionization function.

TABLE 5. SPECTRAL DISTRIBUTION OF FILTERED GAMMA-RAYS

filter cm. lead		0·0	0·2	0·4	1·0	1·4	2·0	3·0	4·0	5·0
λ (X.U.)	μ (lead)	ΔE_0	ΔE_2	ΔE_4	ΔE_{10}	ΔE_{14}	ΔE_{20}	ΔE_{30}	ΔE_{40}	ΔE_{50}
4·15	0·456	1·4	1·3	1·2	0·9	0·76	0·58	0·37	0·23	0·15
4·64	0·465	3·7	3·4	3·1	2·3	1·95	1·47	0·93	0·58	0·29
5·10	0·483	3·9	3·0	2·7	2·0	1·67	1·25	0·77	0·47	0·29
5·57	0·490	7·4	6·7	6·1	4·6	3·74	2·79	1·71	1·05	0·64
5·91	0·502	6·5	5·9	5·3	3·9	3·21	2·37	1·44	0·87	0·53
6·94	0·534	30·3	27·3	24·0	17·8	14·10	10·40	6·11	3·58	2·10
8·89	0·619	22·7	20·1	17·7	12·2	9·50	6·59	3·55	1·92	1·03
9·89	0·670	7·7	6·7	5·9	4·0	3·02	2·02	1·04	0·53	0·27
10·92	0·724	8·6	7·4	6·4	4·2	3·10	2·01	0·98	0·47	0·23
13·1	0·832	9·1	7·6	6·5	3·9	2·83	1·72	0·75	0·32	0·14
16·0	1·060	6·0	4·8	3·9	2·1	1·35	0·72	0·25	0·09	0·03
20·2	1·384	21·9	16·7	12·6	5·5	3·16	1·38	0·35	0·09	0·02
34·9	3·31	7·1	4·1	1·9	0·26	0·07	—	—	—	—
41·6	4·52	10·4	4·2	1·7	0·11	0·02	—	—	—	—
50·8	6·60	13·9	3·7	1·0	0·02	—	—	—	—	—
	$\Sigma\Delta E$	160	123	100	63·8	48·5	33·3	18·3	10·2	5·7

For photoelectric absorption, the Hulme formula indicates that the coefficient τ is roughly proportional to $\lambda^{2\cdot1}$ over these wave-lengths. Thus in a material in which the absorption was purely photoelectric

$$\lambda_{\text{eff.}} = \sqrt[2\cdot1]{\left(\frac{\Sigma\lambda^{2\cdot1}\Delta E}{\Sigma\Delta E}\right)}.$$

These effective wave-lengths are plotted in curve B of figure 4. They are considerably higher than those for scattering, particularly at the lower filtrations. It is clear that the effective wave-lengths for absorption in lead lie between curves A and B. A weighted mean of the two is required. The method adopted is, for each filtration, to weight the effective wave-length for photoelectric absorption by the proportion of the total energy which is absorbed in this way, and similarly with the wave-length for scattering. Thus

$$\lambda_{\text{eff.}} = \frac{{}_p\lambda_{\text{eff.}}\,\Sigma\frac{\tau}{\mu}\,\Delta E + {}_s\lambda_{\text{eff.}}\,\Sigma\frac{\sigma}{\mu}\,\Delta E}{\Sigma\Delta E},$$

where ${}_p\lambda_{\text{eff.}}$ is the effective λ for photoelectric absorption, and ${}_s\lambda_{\text{eff.}}$ is the effective λ for absorption by scattering. Nuclear absorption has been neglected. The wave-lengths obtained in this way are shown in curve C of figure 4, and these have been adopted as the most probable values for the radiations used in the experiments. They do not, in fact, differ greatly from the simple mean values mentioned at the beginning of this section. Identical calculations, using the gamma-ray spectrum of

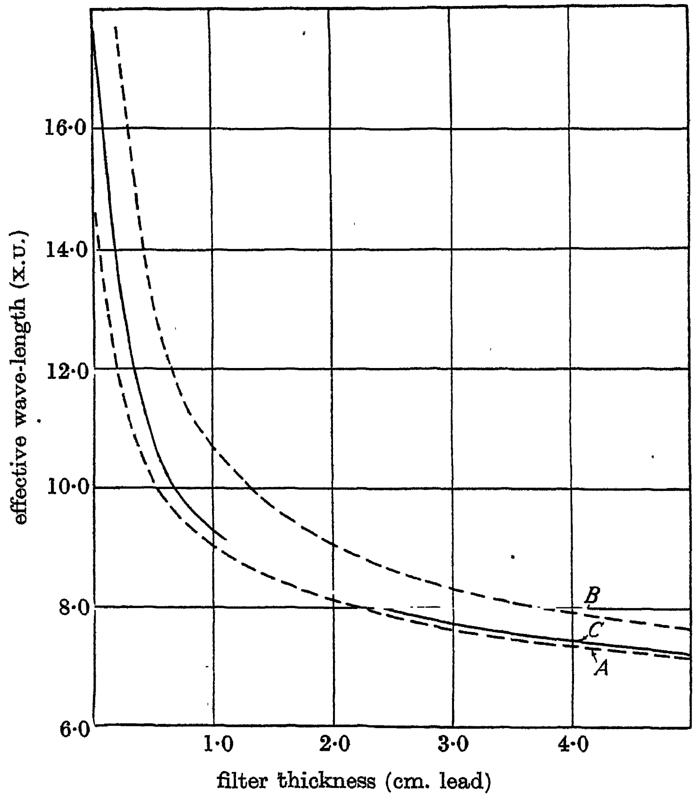


FIGURE 4. Effective wave-length of filtered gamma-rays from radium (B+C) (Skobelzyn spectrum). Curve A, scattering absorption, ${}_s\lambda_{\text{eff.}} = \left(\frac{\Sigma\lambda^{\frac{1}{2}}\Delta E}{\Sigma\Delta E}\right)^2$. Curve B, photoelectric absorption, ${}_p\lambda_{\text{eff.}} = \sqrt[2\cdot1]{\left(\frac{\Sigma\lambda^{2\cdot1}\Delta E}{\Sigma\Delta E}\right)}$. Curve C, weighted mean for absorption in lead,

$$\lambda_{\text{eff.}} = \frac{{}_p\lambda_{\text{eff.}}\,\Sigma\frac{\tau}{\mu}\,\Delta E + {}_s\lambda_{\text{eff.}}\,\Sigma\frac{\sigma}{\mu}\,\Delta E}{\Sigma\Delta E}.$$

Ellis & Aston, lead to effective wave-lengths which are greater by about 1 x.u. for the same filtration over most of the range.

In connexion with the above calculation, it is of interest to observe the variation of the relative proportions of photoelectric and scattering absorption in lead as the primary radiation is more heavily filtered. This is shown in table 6.

TABLE 6. PHOTO- AND SCATTERING ABSORPTION IN LEAD

filter cm. lead	0·0	0·2	0·4	1·0	1·4	2·0	3·0	4·0	5·0
% photo-absorption	25·4	29·2	25·8	21·8	20·7	19·5	18·3	17·5	16·9
% scattering absorption	64·6	70·8	74·2	78·2	79·3	80·5	81·7	82·5	83·1

Comparison of theory and experiment

The filtration scale of figure 3*a* can now be converted into a scale of wave-lengths, estimated from the Skobelzyn spectrum, with the results shown in figure 5*a*, curve *B*. The experimental values for the total mass absorption coefficients in lead are plotted for comparison with those calculated on theoretical grounds in a previous section.

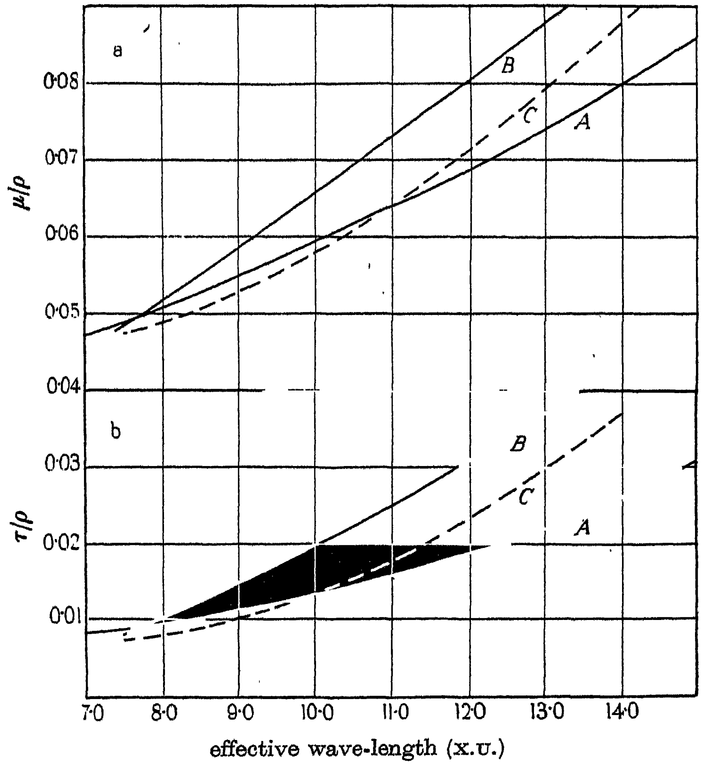


FIGURE 5. *a*, Mass absorption coefficients of gamma-rays in lead. *b*, Mass photoelectric absorption coefficients of gamma-rays in lead. Curves *A*, calculated from theory. Curves *B*, experimental (wave-lengths from Skobelzyn spectrum). Curves *C*, experimental (wave-lengths from Ellis & Aston spectrum).

Although there is fair agreement in the absolute values of μ/ρ at short wave-lengths, the experimental coefficients tend to rise more rapidly with increasing wave-length than the theoretical values. The Klein-Nishina values for the scattering coefficient σ/ρ have been well established experimentally, and nuclear absorption is negligible above 10 X.U. The photoelectric coefficients have therefore been separated from the rest, and plotted in figure 5*b*. Here the theoretical values are from Hulme *et al.* (1935), and the experimental figures have been obtained by subtracting the calculated $(\sigma/\rho+\chi/\rho)$ from the measured total coefficient μ/ρ. As would be expected, the same tendency is apparent, the experimental values of τ/ρ rising more rapidly with increasing wave-length than the predictions of theory. It is of interest to compare the variations of photoelectric absorption for gamma-rays with that observed in the X-ray region. An analysis of the experimental results shows that they are represented approximately by the formula

$$\frac{\tau}{\rho} = \text{const.} \times \lambda^{2\cdot 8}.$$

The exponent of λ in this formula is in better agreement with the value 2·92 normally found for X-rays than with the more generally accepted 2·2 for gamma-rays, but the proportionality constant is considerably larger. The value 2·8 for the exponent, however, should be accepted with some reserve, as the range of wave-lengths observed in the present experiments is very limited.

There is some evidence that the Skobelzyn results on the gamma-ray spectrum tend to overestimate the energy content of the shorter wave-length components. This would lead to effective wave-lengths for the heterogeneous beam which are too small. It is unlikely that this can fully account for the difference between theory and experiment shown in figure 5. Although, as pointed out earlier, the method is not entirely satisfactory, the effective wave-lengths were also calculated using the spectral distribution obtained by the method of Ellis & Aston (1930), which is insensitive to the shorter wave-lengths. Using these figures (figure 5, curves *C*), the absolute discrepancy between theory and experiment is somewhat less, though still large. The variation of τ/ρ with wave-length is almost identical whichever spectrum is used, the exponent of λ in the equation above again being about 2·8.

I wish to thank Miss D. R. Tavener for her assistance in the somewhat tedious computations involved. Thanks are also due to the Medical Research Council for the loan of part of the radium used in the experiments.

REFERENCES

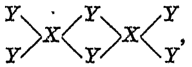
Ahmad 1924*a* *Proc. Roy. Soc.* A, **105**, 507.
Ahmad (& Stoner) 1924*b* *Proc. Roy. Soc.* A, **106**, 8.
Ahmad 1925 *Proc. Roy. Soc.* A, **109**, 207.
Bethe & Heitler 1934 *Proc. Roy. Soc.* A, **146**, 83.
Ellis & Aston 1930 *Proc. Roy. Soc.* A, **129**, 180.

Glasser 1928 *Strahlentherapie*, **29**, 549.
Gray, J. A. 1927 *Trans. Roy. Soc. Can.* **21**, 163.
Gray, L. H. 1931 *Proc. Camb. Phil. Soc.* **27**, 103.
Gray, L. H. 1936 *Proc. Roy. Soc.* A, **156**, 578.
Hulme, McDougall, Buckingham & Fowler 1935 *Proc. Roy. Soc.* A, **149**, 131.
Klein & Nishina 1928 *Z. Phys.* **52**, 853.
Kohlrausch 1917 *S.B. Akad. Wiss. Wien*, **126**, 441.
Mayneord 1933 *Brit. J. Radiol.* **6**, 598.
Mayneord & Roberts 1937 *Brit. J. Radiol.* **10**, 365.
Read & Lauritsen 1936 *Phys. Rev.* **45**, 433.
Rutherford, Chadwick & Ellis 1930 *Radiations from Radioactive Substances*, pp. 466–80. Camb. Univ. Press.
Sievert 1932 *Acta Radiol., Stockh.*, Suppl. 14.
Sizoo & Williams 1938 *Physica, Eindhoven*, **5**, 100.
Skobelzyn 1927 *Z. Phys.* **43**, 354.
Skobelzyn 1929 *Z. Phys.* **58**, 595.
Stoner 1929 *Phil. Mag.* **7**, 841.
Tandberg 1937 Uppsala Dissertation.
Tarrant 1932 *Proc. Camb. Phil. Soc.* **28**, 475.

The normal vibrations of bridged X_2Y_6 molecules

By R. P. Bell and H. C. Longuet-Higgins

(*Communicated by E. J. Bowen, F.R.S.—Received* 7 *February* 1944)

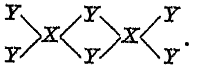
The molecular model considered is , with a tetrahedral distribution of valencies round the atom X. This structure applies to the dimeric trihalides of aluminium, indium and gallium, and to the hydrides B_2H_6 and Ga_2H_6. Symmetry classes and selection rules are given, followed by equations for the normal vibration frequencies derived from a valency-force treatment, taking into account the interactions between unlinked atoms in the ring. A special treatment is necessary for the bending vibration in which the potential energy is proportional to the fourth power of the displacement.

The observed infra-red and Raman spectra of diborane can be satisfactorily assigned on the basis of this model, fifteen out of the seventeen active fundamentals being observed. These fifteen frequencies are predicted with an average error of 2 % and a maximum error of 6 % by the insertion of six adjustable force constants in the theoretical equations. The observed and calculated vibration frequencies are used to calculate the specific heat of diborane in the range 100–300° K. The calculated values agree with experiment, except for a small discrepancy at the highest temperature which can be reasonably accounted for.

The above results provide strong evidence that diborane has a bridge configuration rather than one resembling ethane. The force-constants obtained show that the terminal B-H links are normal single bonds, while the bonds forming the bridge are considerably weaker. This is in harmony with the greater length of the bridge links, and the small number of electrons available for their formation. It is concluded that the hypothetical reaction $2BH_3 \rightarrow B_2H_6$ involves the absorption of at least 15 kcal./mol.

A simplified form of the frequency equations is used to calculate the Raman frequencies of the dimeric halides of aluminium. The results are in semi-quantitative agreement with the observed frequencies, and lead to an assignment which is consistent with the observed intensities.

Introduction

The only vibrational analysis so far carried out for molecules of formula X_2Y_6 relates to the ethane structure $Y_3X\,.\,XY_3$, with or without free rotation of the XY_3 groups. This model will apply to ethane itself, to Si_2H_6, and to the hexa-halogen derivatives of these compounds. There is, however, an alternative structure for a molecule X_2Y_6, in which two of the Y atoms form a bridge between the X atoms, i.e. . There is direct evidence that this structure is the correct one for the dimeric trihalides of aluminium, indium and gallium (Palmer & Elliott 1938; Brode 1940), and it probably also applies to other dimeric halides such as Fe_2Cl_6 and Au_2Cl_6. Moreover, it has been recently shown (Longuet-Higgins & Bell 1943) that diborane, B_2H_6, and the recently discovered Ga_2H_6 (Wiberg & Johannsen 1942) probably also have a bridged structure. Experimental data are available for the Raman spectra of the aluminium halides, and for both Raman and infra-red spectra of diborane. It therefore seems worth while to carry out a quantitative treatment of the vibrations of the bridged molecule.

SYMMETRY PROPERTIES

In most of the molecules mentioned above, the valencies of the X atoms are probably tetrahedrally disposed, so that the molecule X_2Y_6 is non-planar. In a few cases (e.g. Au_2Cl_6) the valencies, and hence the whole molecule, are probably planar. In either case the molecule belongs to the symmetry group D_{2h}, like ethylene. Since the ethane structure has symmetry D_{3h} or D_{3d} the vibrational spectrum should provide a useful criterion for deciding between the ethane and the bridge structures.

In the present paper we shall deal only with the non-planar model. The symmetry characteristics of the molecule are then conveniently taken as three planes at right angles: σ_x through the two central Y atoms and perpendicular to the line joining the X atoms, σ_y containing the two X atoms and the four terminal Y atoms, and σ_z containing the two X atoms and the two central Y atoms. The symmetry classes, selection rules and number of vibrations in each class are given in table 1, the symbols used being those of Placzek (1934). It will be seen that there are no degenerate frequencies, and that no vibrations are active both in the Raman and in the infra-red spectrum. In both these points there is a contrast with the ethane model of symmetry D_{3h} or D_{3d}.

Figure 1 illustrates approximately the character of the vibrations listed in table 1, assuming that the bending force constants are considerably smaller than the stretching ones. The ring is perpendicular to the plane of the paper, and the signs + and − represent motion out of and into this plane respectively. No attempt has been made to show all the motions of any given vibration.

TABLE 1

	symmetry with respect to				
class	σ_x	σ_y	σ_z	activity	vibrations
A_{1g}	*s*	*s*	*s*	Raman (*p*)	ν_1 ν_2 ν_3 ν_4
A_{1u}	*as*	*as*	*as*	inactive	ν_5
B_{1g}	*as*	*as*	*s*	Raman (*dp*)	ν_6 ν_7
B_{1u}	*s*	*s*	*as*	infra-red (M_z)	ν_8 ν_9 ν_{10}
B_{2g}	*as*	*s*	*as*	Raman (*dp*)	ν_{11} ν_{12}
B_{2u}	*s*	*as*	*s*	infra-red (M_y)	ν_{13} ν_{14}
B_{3g}	*s*	*as*	*as*	Raman (*dp*)	ν_{15}
B_{3u}	*as*	*s*	*s*	infra-red (M_x)	ν_{16} ν_{17} ν_{18}

VALENCY-FORCE TREATMENT

The nomenclature used for the dimensions of the molecule at rest is shown in figure 2. M and m represent the masses of X and Y respectively, and it is convenient to distinguish the central Y atoms by the symbol Y'. The expression assumed for the potential energy is

$$\begin{aligned} V = {} & 4\,.\,\tfrac{1}{2}f_1(\Delta XY')^2+4\,.\,\tfrac{1}{2}f_2(\Delta XY)^2 \\ & +2\,.\,\tfrac{1}{2}d_1a^2(\Delta Y'\widehat{X}Y')^2+2\,.\,\tfrac{1}{2}d_2b^2(\Delta Y\widehat{X}Y)^2+2\,.\,\tfrac{1}{2}d_3a^2(\Delta X\widehat{Y'}X)^2 \\ & +8\,.\,\tfrac{1}{2}d_4ab(\Delta Y\widehat{X}Y')^2+\tfrac{1}{2}g_1(\Delta XX)^2+\tfrac{1}{2}g_2(\Delta Y'Y')^2, \end{aligned} \qquad (1)$$

which will apply to any of the normal vibrations listed in table 1. Of the force constants in equation (1), f_1 and f_2 refer to bond-stretching, there being two different constants because the linkage between X and Y in the ring differs in nature from the terminal linkages. Similarly, there are three different bending force constants (d_1, d_2 and d_4) for valencies attached to X, and one (d_3) for valencies attached to Y'.

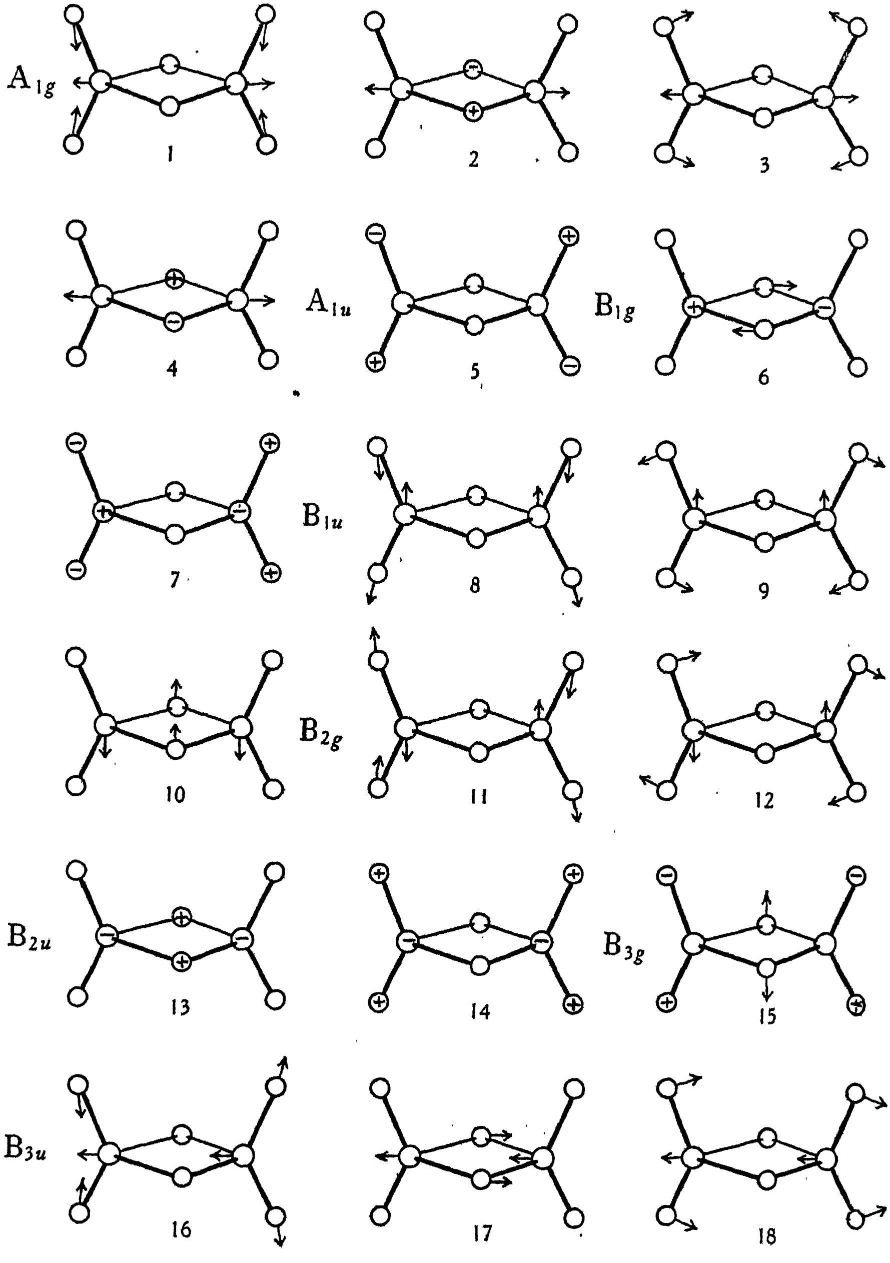


FIGURE 1

These six constants are all that is required in a simple valency-force treatment for the bonds shown in figure 2, but we have added the terms in g_1 and g_2, corresponding to interactions between the pairs of opposite atoms in the ring. Although the ordinary valency picture represents these atoms as unlinked, they are likely to be much closer together than is common for unlinked atoms, and their interactions may be considerable. These terms only enter into vibrations of the class A_{1g}, since in all other classes the distances XX and $Y'Y'$ remain unchanged.

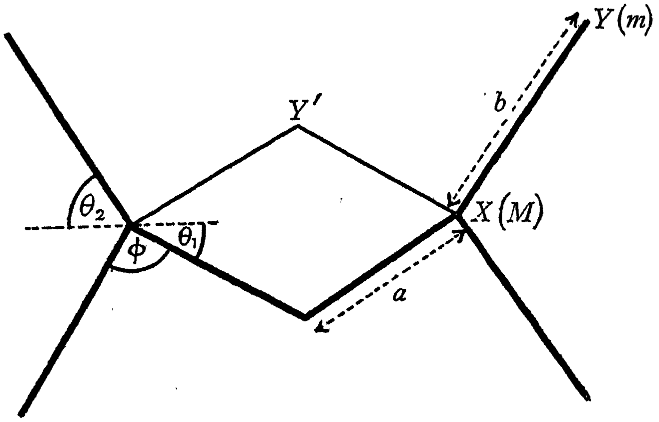


FIGURE 2

In deriving the equations for the normal vibration frequencies we have used two sets of co-ordinates: (*a*) the Cartesian co-ordinates of the atoms, and (*b*) the bond lengths and angles. In the absence of degeneracy either set are satisfactory symmetry co-ordinates, and the secular equation factorizes so that each symmetry class can be treated separately. The equation for each class has been derived independently in both sets of co-ordinates, agreement between the two methods constituting a valuable check on the rather involved algebra.

It is convenient to introduce the following abbreviations:

$$\left.\begin{aligned}
&c_1 = \cos\theta_1, \quad c_2 = \cos\theta_2, \quad s_1 = \sin\theta_1, \quad s_2 = \sin\theta_2,\\
&\frac{1}{m_1} = \frac{1}{m}+\frac{1}{M}, \quad \frac{1}{m_2} = \frac{1}{m}+\frac{2}{M}, \quad \frac{1}{m_3} = \frac{1}{m}+\frac{3}{M}, \quad \frac{1}{m_4} = \frac{c_1^2}{m}+\frac{s_1^2}{M},\\
&\frac{1}{m_5} = \frac{s_1^2}{m}+\frac{c_1^2}{M}, \quad \frac{1}{m_6} = \frac{1}{m}+\frac{2c_2^2}{M}, \quad \frac{1}{m_7} = \frac{1}{m}+\frac{2s_2^2}{M}, \quad \frac{1}{m_8} = \frac{4c_1c_2}{M},\\
&\mu = 4\pi^2\nu^2, \quad d_5 = 2d_4/(1-c_1^2c_2^2), \quad D = s_1^2f_1+2c_1^2d_3, \quad E = c_1^2f_1+2s_1^2d_1,\\
&K_1 = f_1+c_1^2g_1+s_1^2g_2, \quad K_2 = f_2, \quad K_3 = 2d_1+2d_3+\frac{b}{a}s_1^2c_2^2d_5+s_1^2g_1+c_1^2g_2,\\
&K_4 = 2d_2+\frac{a}{b}c_1^2s_2^2d_5, \quad K_5 = s_1c_1s_2c_2d_5, \quad K_6 = (g_1-g_2)\,s_1c_1.
\end{aligned}\right\} \quad (2)$$

The equations for the frequencies of the normal vibrations are then as follows:

$$\left.\begin{aligned}
A_{1g}\quad & \mu^4-\mu^3\Bigl[\frac{2K_1}{m_5}+\frac{K_2}{m_6}+\frac{2K_3}{m_4}+\frac{K_4}{m_7}-4K_5\frac{s_1s_2}{M}-4K_6s_1c_1\Bigl(\frac{1}{m}-\frac{1}{M}\Bigr)\Bigr]\\
&+\frac{\mu^2}{m}\Bigl[2K_1K_2\Bigl(\frac{s_1^2}{m_6}+\frac{c_1^2}{M}\Bigr)+(K_1K_3-K_6^2)\frac{4}{M}+2K_1K_4\Bigl(\frac{s_1^2}{m_7}+\frac{c_1^2}{M}\Bigr)\\
&+2K_2K_3\Bigl(\frac{c_1^2}{m_6}+\frac{s_1^2}{M}\Bigr)+\frac{K_2K_4}{m_2}+2(K_3K_4-K_5^2)\Bigl(\frac{c_1^2}{m_7}+\frac{s_1^2}{M}\Bigr)\\
&-4K_5(2K_1+K_2)\frac{s_1s_2}{M}-4K_2K_6s_1c_1\Bigl(\frac{1}{m_1}-\frac{2s_2^2}{M}\Bigr)-4K_4K_6s_1c_1\Bigl(\frac{1}{m_1}-\frac{2c_2^2}{M}\Bigr)\\
&+8K_5K_6\frac{c_1s_2}{M}\Bigr]-\frac{\mu}{m^2}\Bigl[\frac{4K_2}{M}(K_1K_3-K_6^2)+\frac{4K_1}{M}(K_3K_4-K_5^2)\\
&+2K_2(K_3K_4-K_5^2)\Bigl(\frac{c_1^2}{m_1}+\frac{1}{M}\Bigr)+2K_1K_2K_4\Bigl(\frac{s_1^2}{m_1}+\frac{1}{M}\Bigr)-8K_1K_2K_5\frac{s_1s_2}{M}\\
&-4K_2K_4K_6\frac{s_1c_1}{m_1}+8K_2K_5K_6\frac{c_1s_2}{M}-\frac{4K_4K_6^2}{M}\Bigr]\\
&+\frac{4K_2}{m^3M}[K_1K_3K_4-K_1K_5^2-K_4K_6^2]=0,\\[1ex]
A_{1u}\quad & \mu=\frac{1}{m}\frac{a}{b}s_1^2d_5,\\[1ex]
B_{1g}\quad & \mu^2-\mu\Bigl[\frac{2f_1}{m_4}+d_5s_1^2\Bigl\{2\frac{b}{a}\frac{c_2^2}{m_5}+\frac{a}{b}\frac{1}{m_2}+\frac{1}{m_8}\Bigr\}\Bigr]\\
&\qquad+\frac{2}{m}f_1d_5s_1^2\Bigl[2\frac{b}{a}\frac{c_2^2}{M}+\frac{a}{b}\Bigl(\frac{c_1^2}{m_1}+\frac{1}{M}\Bigr)+\frac{1}{m_8}\Bigr]=0,\\[1ex]
B_{1u}\quad & \mu=0\quad\text{or}\quad\mu^2-\mu\Bigl[\frac{f_2}{m_7}+s_2^2d_5\Bigl\{2\frac{b}{a}\frac{1}{m_1}+\frac{a}{b}\frac{c_1^2}{m_6}+\frac{1}{m_8}\Bigr\}\Bigr]\\
&\qquad+\frac{1}{m}f_2d_5s_2^2\Bigl[2\frac{b}{a}\Bigl(\frac{1}{m_7}+\frac{1}{M}\Bigr)+\frac{a}{b}\frac{c_1^2}{m_2}+\frac{1}{m_8}\Bigr]=0,\\[1ex]
B_{2g}\quad & \mu^2-\mu\Bigl[\frac{f_2}{m_7}+d_5s_2^2\Bigl\{\frac{b}{a}\frac{2}{M}+\frac{a}{b}\frac{c_1^2}{m_6}+\frac{1}{m_8}\Bigr\}\Bigr]+\frac{1}{m}f_2d_5s_2^2\Bigl[\frac{b}{a}\frac{2}{M}+\frac{a}{b}\frac{c_1^2}{m_2}+\frac{1}{m_8}\Bigr]=0,\\[1ex]
B_{2u}\quad & \mu^2-\mu\Bigl[\frac{2D}{m_1}+d_5s_1^2\Bigl\{2\frac{b}{a}\frac{c_1^2c_2^2}{m_1}+\frac{a}{b}\frac{1}{m_2}+\frac{1}{m_8}\Bigr\}\Bigr]+\frac{2Dd_5}{mm_3}\frac{a}{b}s_1^2=0,\\[1ex]
B_{3g}\quad & \mu=\frac{d_5}{m}\Bigl\{2\frac{b}{a}s_2^2+\frac{a}{b}s_1^2\Bigr\},\\[1ex]
B_{3u}\quad & \mu^3-\mu^2\Bigl[\frac{2E}{m_1}+\frac{f_2}{m_6}+\frac{2d_2}{m_7}+d_5\Bigl\{2\frac{b}{a}\frac{s_1^4c_2^2}{m_1}+\frac{a}{b}\frac{c_1^2s_2^2}{m_7}-\frac{s_1^2s_2^2}{m_8}\Bigr\}\Bigr]+\frac{\mu}{m}\Bigl[2E\Bigl\{f_2\Bigl(\frac{1}{m_6}+\frac{1}{M}\Bigr)\\
&+2d_2\Bigl(\frac{1}{m_7}+\frac{1}{M}\Bigr)\Bigr\}+2d_5\Bigl\{2\frac{b}{a}s_1^4c_2^2d_2+\frac{a}{b}c_1^2s_2^2E\Bigr\}\Bigl(\frac{1}{m_7}+\frac{1}{M}\Bigr)\\
&+\frac{2f_2d_2}{m_2}+f_2d_5\Bigl\{2\frac{b}{a}s_1^4c_2^2\Bigl(\frac{1}{m_6}+\frac{1}{M}\Bigr)+\frac{a}{b}\frac{c_1^2s_2^2}{m_2}-\frac{s_1^2s_2^2}{m_8}\Bigr\}\Bigr]\\
&-\frac{2f_2}{m^2m_3}\Bigl[2d_2d_5\frac{b}{a}s_1^4c_2^2+d_5E\frac{a}{b}c_1^2s_2^2+2d_2E\Bigr]=0.
\end{aligned}\right\}\qquad(3)$$

Kohlrausch & Wagner (1942) have recently given expressions for the Raman frequencies of a simplified type of bridged molecule, and a later paper (Wagner 1943) gives corrections to these expressions and extends the treatment to the infrared frequencies. They consider a molecule in which all the bond lengths are equal and the valencies of the atom X are directed towards the vertices of a regular tetrahedron. Their expressions should therefore be obtainable from ours (allowing for differences of notation) by putting $a = b$, $2\theta_1 = 2\theta_2 = 109^\circ$. In most cases there is agreement, but we believe their expressions for B_{2g} and B_{3u} to be in error. The correct equation for B_{2g} is (in their notation)

$$B_{2g} \quad \mu^2 - \mu\left\{\left(\frac{1}{m}+\frac{4}{3M}\right)f_2 + \frac{1}{2}\left(\frac{1}{m}+\frac{32}{3M}\right)d_3\right\} + \frac{1}{2m}\left(\frac{1}{m}+\frac{12}{M}\right)f_2 d_3 = 0.$$

In the case of B_{3u} their equation involves a fundamental error, since it does not contain the bending force constant for the angle $Y\widehat{X}Y$ (d_2 in both notations). Kohlrausch & Wagner have not given explicitly the equation for class A_{1g}, but we agree with their (corrected) expression for the secular determinant.

THE FOURTH-POWER VIBRATION

In the above treatment the frequency ν_{10} of class B_{1u} comes out as zero, since for a first-order infinitesimal displacement the angles and the distances undergo only second-order changes. It has been shown elsewhere (Bell 1944) that in this type of ring-bending the potential energy of the system is proportional to the fourth power of the displacement, and the theory of such 'fourth-power vibrations' has been considered. For a single particle of mass m having a potential energy given by $V(x) = Ax^4$ it was shown that the lowest spectrum frequency ν_0 is given by

$$\nu_0 = 0{\cdot}163(hA/m^2)^{\frac{1}{3}}. \tag{4}$$

This equation can be used to calculate the frequency ν_{10} in the molecule X_2Y_6, provided that some simplifying assumptions are made. In the first place, it is necessary to treat the fourth-power vibration as a separate problem, divorced from the other vibrations of the molecule. Secondly, the terminal XY_2 groups are considered to move as a whole without distortion. There is no strict justification for either of these simplifications, but they are both physically reasonable provided that the fourth-power frequency is considerably lower than the other frequencies of the system: this is likely to be so in practice. If the XY_2 groups undergo an arbitrary small displacement z perpendicular to the plane of the ring, then the vibration of the ring can be described by the following scheme of displacements:

	δx	δy	δz
XY_2	$\pm\frac{\lambda_1}{a}z^2$	0	$+z$
Y'	0	$\pm\frac{\lambda_2 q^2}{a}z^2$	$-qz$

(5)

where $q = (M+2m)/m$, and λ_1 and λ_2 are parameters to be determined subsequently. The changes in distances and angles for small z can then be obtained from the ordinary geometrical expressions by expanding in powers of z^2 and rejecting higher powers. The results are:

$$\left.\begin{aligned}
\varDelta XY' &= \frac{z^2}{2a}\{(1+q)^2-2c_1\lambda_1-2q^2s_1\lambda_2\},\\
\varDelta Y'\widehat{X}Y' &= -\frac{z^2}{a^2c_1}\{(1+q)^2s_1-2c_1s_1\lambda_1+2q^2c_1^2\lambda_2\},\\
\varDelta X\widehat{Y'}X &= -\frac{z^2}{a^2s_1}\{(1+q)^2c_1+2s_1^2\lambda_1-2q^2c_1s_1\lambda_2\},\\
\varDelta Y\widehat{X}Y' &= -\frac{s_1c_2}{2(1-c_1^2c_2^2)^{\frac{1}{2}}}\varDelta Y'\widehat{X}Y',\quad \varDelta XX = -\frac{2\lambda_1}{a}z^2,\quad \varDelta Y'Y' = -\frac{2q^2\lambda_2}{a}z^2.
\end{aligned}\right\}\qquad(6)$$

In terms of the force constants already defined, the potential energy is then given by

$$\begin{aligned}
V &= 4\,.\,\tfrac{1}{2}f_1(\varDelta XY')^2+2\,.\,\tfrac{1}{2}d_1a^2(\varDelta Y'\widehat{X}Y')^2+2\,.\,\tfrac{1}{2}d_3a^2(\varDelta X\widehat{Y'}X)^2\\
&\quad+8\,.\,\tfrac{1}{2}d_4ab(\varDelta Y\widehat{X}Y')^2+\tfrac{1}{2}g_1(\varDelta XX)^2+\tfrac{1}{2}g_2(\varDelta Y'Y')^2\\
&= \frac{1}{2}\frac{z^4}{a^2}\Big[f_1\{(1+q)^2-2c_1\lambda_1-2q^2s_1\lambda_2\}^2\\
&\quad+\left(\frac{2d_1}{c_1^2}+\frac{b}{a}\frac{s_1^2c_2^2}{c_1^2}d_5\right)\{(1+q)^2s_1-2c_1s_1\lambda_1+2q^2c_1^2\lambda_2\}^2\\
&\quad+\frac{2d_3}{s_1^2}\{(1+q)^2c_1+2s_1^2\lambda_1-2q^2c_1s_1\lambda_2\}^2+4g_1\lambda_1^2+4g_2q^4\lambda_2^2\Big]\\
&= Az^4,
\end{aligned}\qquad(7)$$

whence λ_1 and λ_2 are determined by minimizing V, i.e. putting $\dfrac{\partial V}{\partial\lambda_1} = \dfrac{\partial V}{\partial\lambda_2} = 0$. The kinetic energy is to the same approximation

$$T = 2\,.\,\tfrac{1}{2}m(q\dot{z})^2+2\,.\,\tfrac{1}{2}qm\dot{z}^2 = \tfrac{1}{2}\,.\,2mq(1+q)\dot{z}^2.\qquad(8)$$

Comparing this with equation (4) for a single particle, then

$$\nu_{10} = 0{\cdot}163\left\{\frac{hA}{4m^2q^2(1+q)^2}\right\}^{\frac{1}{3}},\qquad(9)$$

where A is defined by (7) in terms of known quantities. It should be noted that ν_{10} no longer belongs strictly to the symmetry class B_{1u}, and its activity is thus not restricted to the infra-red spectrum. However, its Raman activity is likely to be weak, since the relevant changes in distances and angles are of the second order, compared with the first-order change in electric moment.

APPLICATION TO THE SPECTRUM OF DIBORANE

Experimental data. The Raman spectrum of liquid diborane has been studied by Anderson & Burg (1938), and the infra-red spectrum of the gas by Stitt (1941). The results obtained are given in the first three columns of table 2. (Two of the Raman shifts originally recorded were later attributed to mercury lines.) In the infra-red spectrum the low resolving power used makes it difficult to determine the band contours with certainty. However, three of the bands appear to be of the parallel

TABLE 2. INFRA-RED AND RAMAN SPECTRA OF DIBORANE

infra-red spectrum

ν (obs.) cm.$^{-1}$	uncertainty	intensity	assignment
393 412 ∥ (?) 431	3	strong	fundamental ν_{10}
679	3	weak	$1863-1180=683$
831	3	weak	$1608-793=815$
981	2	strong	fundamental ν_{14}
1154 1178 ∥ 1197	2	very strong	fundamental ν_{18}
1292	2	medium	$2489-1178=1311$
1377	3	medium	$2558-1180=1378$
1405	5	medium	fundamental ν_9
1587 1608 ∥ 1626	3	very strong	fundamental ν_{17}
1722	10	weak	?
1863	4	strong	fundamental ν_{13}
1993	6	medium	$1008+981=1989$
			$1178+820=1998$
			$1178+793=1971$
2134	10	very weak	$1180+981=2161$
2217	10	very weak	$1405+820=2225$
			$1405+793=2198$
2353	8	strong	$1180+1178=2358$
2558	10	strong	fundamental ν_{16}
2625	10	strong	fundamental ν_8

Raman spectrum

ν (obs.) cm.$^{-1}$	intensity	assignment
793	5	fundamental ν_4 $(B^{11}-B^{11})$
806	4	fundamental ν_4 $(B^{11}-B^{10})$
821	3 (broad)	fundamental ν_4 $(B^{10}-B^{10})$, ν_{12}
1008	diffuse	fundamental ν_7
1180	2	fundamental ν_3
2102	10	fundamental ν_2
2489	3	fundamental ν_{11}
2523	10	fundamental ν_1

type, and for these the positions of the three branches are given in the table. In other cases only the centres of the bands are recorded. Bands in the range 3000–6000 cm.$^{-1}$ have not been included in the table, since their positions are not known with sufficient accuracy to permit of any certain assignment. They will all be combination frequencies, and all of them can be accounted for in several different ways.

Frequency assignments for the ethane model. Attempts to analyse the vibrational spectrum on this basis (Stitt 1941; Bauer 1942) have not been very successful. The observed spectrum is much more complex than that of ethane, and in particular there are too many strong lines observed in the region corresponding to B-H bond-stretching (2000–3000 cm.$^{-1}$). This difficulty appears in a simple form in the Raman spectrum, where the two most intense frequencies, 2102 and 2523, must both be attributed to totally symmetrical fundamentals, though the ethane model predicts only *one* stretching frequency of this type. Moreover, the infra-red spectrum contains eight bands which appear to be strong enough for fundamentals, compared with five predicted by the ethane model, and of these the one at 412 cm.$^{-1}$ is much too low to be readily accounted for. In order to meet these difficulties Stitt (1941) has postulated the existence of one or more low-lying electronic levels in diborane, and has given two alternative frequency assignments on this basis. However, these assignments still involve difficulties, and there is no independent evidence for the existence of low-lying electronic levels in diborane or in similar molecules. Anderson & Burg (1938) have suggested that the two Raman lines 2102 and 2523 may represent a resonance splitting, but it is most unlikely that a separation as great as 421 cm.$^{-1}$ could arise from this cause.

Frequency assignments for the bridge model. As previously mentioned (Longuet-Higgins & Bell 1943) these difficulties disappear if the bridge formula is adopted for diborane. The absence of degeneracy allows eight distinct frequencies to appear in the infra-red (see table 1), and the low frequency 412 is naturally interpreted as the fourth-power bending vibration ν_{10}. Moreover, there are now two totally symmetrical types of bond-stretching (ν_1 and ν_2) to be allotted to the Raman frequencies 2102 and 2523.

Most of the fundamental frequencies can be allotted with little ambiguity, and some information is given by the infra-red band contours. The bridge structure for B_2H_6 is approximately a symmetrical top, the moments of inertia about the y and z axes differing by less than 10 %, and it may therefore be expected that the bands can be classified as parallel (B_{3u}) and perpendicular (B_{1u} and B_{2u}). The expression given by Gerhard & Dennison (1933) predicts a separation of about 40 cm.$^{-1}$ between the P and R branches, which agrees approximately with the bands marked as parallel in table 2. The apparently parallel character of the band at 412 cm.$^{-1}$ is at variance with its allocation as ν_{10} of B_{1u}, but this region of the spectrum is very near the limit of transparency of the prism used, and the contours are uncertain: moreover, ν_{10} is not an ordinary harmonic vibration, and we shall see later that there are other possible reasons for the structure of this band.

The assignments are most readily made by dividing the normal vibrations into the following categories. (The hydrogen atoms in the ring will be distinguished as H′).

(*a*) *Stretching of terminal* B-H *links* (ν_1, ν_8, ν_{11}, ν_{16}). The vibrational frequency of the spectroscopic molecule BH is 2366 (King 1938), and the B-H stretching frequency in triborine triamine, $B_3N_3H_6$, is 2535 (Crawford & Edsall 1939). We should therefore expect these frequencies to lie in the range 2300–2700, with $\nu_1 \simeq \nu_{16}$ (cf. figure 1). From its intensity $\nu_1 = 2523$, leaving $\nu_{11} = 2489$. In the infra-red ν_{16} ($\simeq \nu_1$) = 2558, leaving 2353 or 2625 for ν_8. Of these two frequencies only the first can be satisfactorily interpreted as a combination tone, so that we have assumed $\nu_8 = 2625$. ν_{16} should be a parallel band, but this would probably escape detection owing to the low resolving power.

(*b*) *Ring-stretching* (ν_2, ν_6, ν_{13}, ν_{17}). These frequencies will be somewhat lower than those of type (*a*), since the linkage B-H′ will be weaker than the terminal B-H link. The intense Raman line 2102 is clearly the totally symmetrical ν_2, and 1608 (parallel) = ν_{17}. This leaves $\nu_{13} = 1861$ in the infra-red, while ν_6 is probably not observed.

(*c*) B-B *vibration* (ν_4). This differs from type (*b*) in that the B-H′ links are only slightly stretched, while the forces between unlinked atoms in the ring come into play. No prediction of frequency can be made, but the vibration should appear strongly in the Raman spectrum and should show a marked separation for the isotopes B^{10} and B^{11}: hence it is identified as the triplet 793, 806, 821. The component at 821 is broad and is considerably more intense than would be expected from the relative abundance of the isotopes, and it is likely that there is a second Raman frequency at this point.

(*d*) *Angular motion of terminal* B-H *groups* (ν_3, ν_5, ν_7, ν_9, ν_{12}, ν_{14}, ν_{15}, ν_{18}). By analogy with other compounds of hydrogen these frequencies should lie in the neighbourhood of 1000. They can be further classified as twisting (ν_5 and ν_{15}), deformation (ν_3 and ν_{18}), rocking (ν_9 and ν_{12}) and bending (ν_7 and ν_{14}) (cf. Linnett 1938). The twisting frequency ν_5 is inactive, and ν_{15} is likely to appear very weakly if at all. ν_3 and ν_{18} both involve changes in the angle $H\hat{B}H$; ν_3 should be the most intense Raman frequency in this region, and ν_{18} a parallel band. Clearly $\nu_3 = 1180$ and $\nu_{18} = 1178$. The remaining frequencies involve only changes in the angle $H\hat{B}H'$, and presumably include 821 and 1008 in the Raman spectrum, and 981 in the infra-red. A closer assignment can only be made by appealing to the equations (3), which show that $\nu_1 > \nu_{12}$ and $\nu_9 > \nu_{14}$, for any reasonable values of the force constants. This suggests $\nu_{12} = 821$, $\nu_7 = 1008$, the diffuse character of the latter being in accord with the considerable isotope effect (*ca.* 20 cm.$^{-1}$) predicted by equation (3) B_{1g}. There are only weak infra-red bands between 412 and 981, so that $\nu_{14} = 981$. The remaining frequency ν_9 might be any one of the bands of medium intensity 1292, 1377 or 1405. In choosing $\nu_9 = 1405$ we have been guided by the results of the quantitative treatment in the next section.

The above assignments account for fifteen out of the seventeen active fundamentals, and one of the two missing frequencies would be expected to have low Raman activity. Moreover, fourteen of the fifteen assignments relate to observed frequencies described as 'strong', and there is only one such observed frequency (2353) which cannot be interpreted as a fundamental. The high intensity of this band may be related to the close coincidence between $\nu_3+\nu_{18} \doteq 2358$, $2\nu_3 = 2360$, $2\nu_{18} = 2356$. The combination tones assigned in table 2 involve only observed fundamentals, and are allowed by the selection rules derived from table 1. An alternative explanation of the weak infra-red frequencies 831 and 2134 is to identify them with the fundamental Raman frequencies 821 and 2102, since in the unsymmetrical molecule $B^{10}B^{11}H_6$ these could appear weakly (and with slightly enhanced frequency) in the infra-red spectrum. Further alternative assignments for combination tones are considered at the end of the next section.

Application of normal co-ordinate treatment. The general equations (3) contain eight force constants, but in the case of diborane it is reasonable to reduce this number to six by putting $d_3 = g_2 = 0$. d_3 refers to the bending of the ring angle $B\widehat{H'}B$, and the two links attached to hydrogen would not be expected to have directional properties. Similarly, g_2 refers to the interaction between the two unlinked H′ atoms in the ring, which is likely to be small compared with the interaction between the boron atoms, represented by g_1. We have also assumed the value $f_2 = 3{\cdot}42 \times 10^5$ dynes/cm. derived by Crawford & Edsall (1939) for B-H stretching in triborine triamine, since the terminal links in diborane should be normal single bonds. The equations also contain the dimensions of the molecule, for which we have taken the values derived from electron-diffraction data (Longuet-Higgins & Bell 1943), i.e. $a = 1{\cdot}39$ A, $b = 1{\cdot}18$ A, $\theta_1 = 50°$, $\theta_2 = 60°$.

There then remain five adjustable force constants to be derived from the observed fundamental frequencies. Of these, the most difficult to fix is d_4 (or the related d_5), which occurs in the equation for every symmetry class. We have chosen the value $d_4 = 0{\cdot}302 \times 10^5$ dynes/cm. as giving the best general representation of the observed bending frequencies. The remaining four force-constants were then derived by assuming successively $\nu_{13} = 1861$, $\nu_{17} = 1608$, $\nu_{18} = 1178$, $\nu_4 = 793$, when all the remaining frequencies could be calculated.

The force constants obtained are given at the head of table 3, followed by the calculated and observed frequencies. The former are given to the nearest 5 cm.$^{-1}$, and those observed frequencies which were used in deriving the force constants are in parentheses. The mean deviation between calculated and observed frequencies is 2 %, and the maximum deviation 6 %. Still better agreement could be obtained by abandoning the assumption $d_3 = g_2 = 0$: for example, if $g_2 \doteq 0{\cdot}3$ the discrepancy of 6 % for ν_2 disappears, and ν_4 is increased by only 2 %. However, it does not seem worth while to make any further adjustments, since the agreement shown in table 3 is quite as good as has been previously obtained for molecules of comparable complexity, using a valency-force potential energy function without cross terms. Moreover, the calculated bending frequencies are rather sensitive to the values taken for the angles θ_1 and θ_2 and the ratio b/a, and it is unlikely that these quantities are given to better than ± 5 % by the electron diffraction data (cf. Diatkina & Sirkin 1942).

TABLE 3. FORCE CONSTANTS AND FREQUENCIES FOR DIBORANE

force-constants (in 10^5 dynes/cm.)

$f_1 = 1{\cdot}43$, $f_2 = 3{\cdot}42$, $d_1 = 0{\cdot}059$, $d_2 = 0{\cdot}243$, $d_3 = 0$, $d_4 = 0{\cdot}302$, $g_1 = 2{\cdot}21$, $g_2 = 0$.

frequencies (in cm.$^{-1}$)

		calc.	obs.	% diff.
A_{1g}	ν_1	2495	2523	−1
	ν_2	1985	2102	−6
	ν_3	1180	1180	0
	ν_4	(792)	792	0
A_{1u}	ν_5	890	(inactive)	—
B_{1g}	ν_6	1540	—	—
	ν_7	1050	1008	+4
B_{1u}	ν_8	2585	2625	−2
	ν_9	1450	1405	+3
	ν_{10}	266, 350, 407	412	−1
B_{2g}	ν_{11}	2580	2489	+4
	ν_{12}	800	821	−2
B_{2u}	ν_{13}	(1861)	1861	0
	ν_{14}	920	981	−6
B_{3g}	ν_{15}	1505	—	—
B_{3u}	ν_{16}	2475	2558	−2
	ν_{17}	(1608)	1608	0
	ν_{18}	(1178)	1178	0

Wagner (1943) has attempted a still greater simplification, putting $a = b$, $2\theta_1 = 2\theta_2 = 109°$, $f_1 = f_2 = 10d_1 = 10d_2 = 10d_3 = 10d_4 = 2{\cdot}5 \times 10^5$, $g_1 = g_2 = 0$, but his calculated frequencies do not agree even qualitatively with the observed spectrum of diborane. This is chiefly because (as already mentioned) his equations for classes B_{2g} and B_{3u} are incorrect, which leads him to identify the observed band at 412 with ν_{18}, and to assume that ν_{10} is not observed.

The fourth-power frequency ν_{10} calls for special comment. When the above values of the force constants are inserted in equations (7) and (9) we find $\nu = 266$ cm.$^{-1}$ for the lowest spectrum frequency, which is too low to be observed directly in the infra-red. It is of course possible that this calculation is considerably in error, but it seems more likely that the observed band centred at 412 is an overtone of this same vibration rather than the fundamental. In a harmonic vibration the transitions $0 \leftrightarrow 1$, $1 \leftrightarrow 2$, $2 \leftrightarrow 3$, etc., all give the same spectrum frequency, but in a fourth-power vibration these frequencies are in the ratio $1 : 1{\cdot}32 : 1{\cdot}53 : 1{\cdot}69 : 1{\cdot}82$, etc. (Bell 1944). If the lowest frequency is in fact 266, this would produce a series of bands at 266, 350, 407, 450, 484, etc., of which the third agrees well with the observed frequency of 412. Moreover, it is easy to show that the intensities of the first two overtones will be comparable with that of the fundamental. In a given mass of gas the integrated intensity of an absorption band for the transition $m \rightarrow n$ is proportional to $(x_{mn})^2 \nu_{mn} \exp(-E_m/kT)$, where E_m is the energy of the mth level, and x_{mn} is the matrix element $\int \psi_m \psi_n x\,dx$ (Bethe 1933). Approximate values of x_{01}, x_{12} and x_{23} can be computed from the wave functions previously derived (Bell 1944), giving $1 : 0{\cdot}69 : 0{\cdot}21$ for the relative intensities of the bands at 266, 350 and 407. Absorption decreases slowly on the short-wave side of the observed band and is still perceptible at 500 cm.$^{-1}$: this can be attributed to the superposition of bands of decreasing intensity corresponding to the transitions 3 → 4, 4 → 5, etc. The apparently parallel nature of the observed band could easily arise from this complex structure in conjunction with the isotope effect.

The details of this interpretation could only be satisfactorily tested by observations with higher resolving power and at lower frequencies. For example, if the lowest frequency were 310 cm.$^{-1}$ instead of 266 cm.$^{-1}$ the observed band at 412 would be the first rather than the second overtone. On the other hand, even if we attach no weight to the calculated value of 266 it does not seem likely that 412 is a fundamental, since such an assumption would predict specific heats which are considerably too low over the whole temperature range (cf. next section).

The calculated values for ν_5, ν_6 and ν_{10} lead to some additional assignments for combination tones, e.g.

$$\begin{array}{rl}
\left.\begin{array}{r} 1008+266^* = 1274 \\ 1540-266^* = 1274 \end{array}\right\} & \text{(obs. 1292)} \\
1608-266^* = 1342 & \text{(obs. 1377)} \\
821+890 \;\; = 1711 & \text{(obs. 1722)} \\
1863+266^* = 2129 & \text{(obs. 2134)} \\
2102+266 \;\; = 2364 & \text{(obs. 2353)}
\end{array}$$

In the cases marked * the selection rule for B_{1u} is violated, but, as already indicated, the fourth-power vibration ν_{10} does not strictly belong to this symmetry class. It may be observed that the assumption $\nu_{10} = 310$ (with the observed 412 as its first overtone) leads to a somewhat more satisfactory interpretation of combination frequencies.

THE SPECIFIC HEAT OF DIBORANE

The specific heat of gaseous diborane has been measured over the temperature range 100–300° K (Stitt 1940). In order to account for the observed values on the basis of the ethane structure it was necessary to assume that the rotation of the —BH_3 groups is restricted by a barrier of either 5000 or *ca.* 15,000 cal./mol., according to the assumptions made about the supposed low-lying electronic states (Stitt 1941). Both figures are higher than the value usually accepted for ethane (*ca.* 3000 cal./mol.), which is surprising, since the hydrogen atoms are further apart in diborane than in ethane, and it is unlikely that the B-B link would have any double-bond character.

It is therefore of interest to use the frequencies in table 3 to calculate the specific heat of the bridge structure. This can be done without introducing any adjustable parameters, since the calculated values can be used for the inactive frequency ν_5 and the unobserved frequencies ν_6 and ν_{15}. The standard harmonic oscillator expression has been used for all the frequencies except the fourth-power vibration ν_{10}. The contribution of this vibration was calculated as previously described (Bell 1944), assuming the lowest spectrum frequency to be 266 cm.$^{-1}$. The values obtained are given in table 4.

The agreement is satisfactory, except at the highest temperature, where the calculated specific heat is somewhat too low. Part of this discrepancy is probably due to neglect of anharmonicity, which would produce too low values for c_v. Anharmonicity is particularly marked in hydrogen compounds and is likely to produce an appreciable correction even at 300° K for the eight frequencies below 1200 cm.$^{-1}$. A correction corresponding to an effective lowering of each of these frequencies by 5 % would be sufficient to remove the discrepancy at 300° K. It is of course also possible that the unobserved twisting frequencies ν_5 and ν_{15} are not correctly predicted by the angular force constant d_4 derived from other types of vibration. If these two frequencies are reduced by about 25 % the observed specific heats are reproduced within experimental error without making any anharmonicity correction.

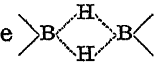
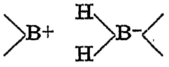
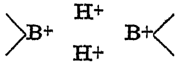
TABLE 4. SPECIFIC HEAT OF DIBORANE

T (°K)	100	150	200	250	300
c_v (cal./mol.) obs.	$6{\cdot}35 \pm 0{\cdot}2$	$6{\cdot}90 \pm 0{\cdot}2$	$8{\cdot}15 \pm 0{\cdot}1$	$9{\cdot}80 \pm 0{\cdot}1$	$11{\cdot}60 \pm 0{\cdot}2$
calc.	6·57	7·15	8·06	9·49	10·80

THE VALENCY STRUCTURE OF DIBORANE

The last two sections can be summed up by saying that the observed vibrational spectrum of diborane is in excellent qualitative and quantitative agreement with the predictions of the bridge model, while the specific heat results are compatible with the same model. This strengthens the evidence previously given (Longuet-Higgins & Bell 1943) in favour of the bridge configuration of diborane. The values of the force-constants given in table 3 also provide material for speculation as to the nature of the links involved in the bridge structure.

The terminal B-H stretching vibrations are well represented by the same force constant f_2 used by Crawford & Edsall (1939) for triborine triamine, thus confirming the assumption that the terminal links are normal single bonds. On the other hand, the constant f_1, referring to the B-H′ link in the bridge, is less than half as great as f_2, as would be expected if the bridge linkages are one-electron or resonance bonds. According to the relation proposed by Douglas Clark (1934) the ratio of the distances B-H and B-H′ should be $(3{\cdot}42/1{\cdot}43)^{\frac{1}{6}} = 1{\cdot}16$, agreeing within experimental error with the electron-diffraction value of $1{\cdot}39/1{\cdot}18 = 1{\cdot}18$. The bending force constants cannot be directly compared with any data, but d_4 and d_2 are close to the values 0·35 and 0·23 given by Crawford & Edsall for planar and non-planar bending in the system H-B-N. On the other hand the constant d_1 is considerably smaller, in harmony with the fact that it refers to the angle between two B-H′ links, which (whatever their exact formulation) contain only two electrons between them.

The value obtained for g_1 (representing interaction between the boron atoms) is an upper limit, since it involves the assumptions $g_2 = d_3 = 0$, but its value cannot be less than about $1{\cdot}5 \times 10^5$ dynes/cm. on any reasonable set of assumptions. In any event, g_1 is less than half as great as would be expected for a B-B link, though it seems rather large to represent an interaction between two unlinked atoms. If the bridge structure  is to be represented as a resonance hybrid of normal valency states, the relatively high value of g_1 could be related to the occurrence of structures such as  among these states. However, it should be noted that in a recent analysis of the vibrational spectrum of cyclo-butane (Wilson 1943) the potential function derived contains large cross-terms which are physically equivalent to interactions between unlinked atoms. The diagonal distance in the cyclo-butane ring is 2·16 A (compared with 1·79 A in diborane), and it would not be generally supposed that the structure of this molecule involves any resonance states with charges on opposite carbon atoms. In our opinion, therefore, the force constants which we have derived do not provide any evidence for the predominance of ionic states in the diborane molecule, as proposed by Sirkin & Diatkina (1941). The whole question is to some extent a formal rather than a real one, since the representation of the bridge linkage as a resonance hybrid is at best an artificial way of representing the actual electron distribution. A better picture of the true state of affairs might be obtained by treating the four-electron system  by the molecular orbital method, which has not yet been attempted.

It is also possible to estimate a lower limit for the energy absorbed in the hypothetical reaction $B_2H_6 \rightarrow 2BH_3$. Experiment shows that although diborane is a very reactive substance at ordinary temperatures and often produces compounds containing the BH_3 group, its vapour density indicates no signs of dissociation into $2BH_3$. It is therefore safe to assume that $p^2_{BH_3}/p_{B_2H_6} < 10^{-3}$ atm. at 300° K, i.e. $\Delta F_{300} > 4{\cdot}2$ kcal./mol. On the other hand, we can estimate the value of ΔF on theoretical grounds. In the molecule B_2H_6 ν_{10} is the only frequency low enough to contribute appreciably to the free energy at room temperature, and in BH_3 the vibrational contribution can presumably be neglected. Thus in the reaction $B_2H_6 \rightarrow 2BH_3$ three translational and three rotational degrees of freedom are formed from six vibrations, five of which contribute nothing to the free energy. The value of ΔF can then be derived from the standard expressions, using the dimensions already given for B_2H_6 and assuming BH_3 to be a planar molecule with B-H = 1·18 A. The result is $\Delta F_{300} - \Delta E_0 = -9{\cdot}4$ kcal./mol., where ΔE_0 is the energy change at absolute zero. Combining this with the experimental result for ΔF we find $\Delta E_0 > 13{\cdot}6$ kcal./mol. as a lower limit for the energy of dissociation of B_2H_6 into $2BH_3$.

THE RAMAN SPECTRA OF THE ALUMINIUM HALIDES

These spectra have been measured in the fused state by Gerding & Smit (1941) and by Rosenbaum (1940). Their data are given in table 5. Since the infra-red frequencies are not available, it is impossible to carry out such a detailed analysis as for diborane, but a reasonably certain assignment can be made by using a simplified form of the valency-force equations. We have calculated the Raman frequencies of classes A_{1g}, B_{1g}, B_{2g} and B_{3g} from equations (3) assuming throughout

$$a = b, \quad 2\theta_1 = 2\theta_2 = 109^\circ,$$

$$f_1 = f_2 = 5g_1 = 5g_2 = 10d_1 = 10d_2 = 10d_4 = 1{\cdot}5 \times 10^5, \quad d_3 = 0.$$

The values obtained are given in table 6, together with the suggested assignments of the observed frequencies. There is of course no exact agreement between the observed and calculated frequencies, but the general pattern of the observed spectra is well reproduced by the simplified equations. The trend of the values shows that better agreement would be obtained if the force constants were assumed to decrease in passing along the series Al_2Cl_6, Al_2Br_6, Al_2I_6; this is what would be expected on chemical grounds.

TABLE 5. OBSERVED RAMAN FREQUENCIES FOR THE ALUMINIUM HALIDES

(Figures in parentheses represent intensity and state of polarization)

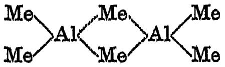
Al_2Cl_6		Al_2Br_6		Al_2I_6	
112	(6, *dp*?)	73	(6)	53	(6)
164	(3, *dp*)	112	(3)	94	(6)
217	(5, *p*)	140	(5)	146	(10)
284	(2, *dp*)	176	(2)	195	(½)
340	(10, *p*)	204	(10)	344	(2½)
438	(½, *dp*?)	221	(½)	406	(3)
506	(3, *p*)	291	(½)		
606	(2½, *dp*)	407	(2½)		
		491	(3)		

TABLE 6. ASSIGNMENTS AND CALCULATED RAMAN FREQUENCIES FOR THE ALUMINIUM HALIDES

		Al_2Cl_6		Al_2Br_6		Al_2I_6	
		calc.	obs.	calc.	obs.	calc.	obs.
A_{1g}	ν_1	519	606	485	491	476	406
	ν_2	341	340	231	204	188	146
	ν_3	200	217	144	140	107	94
	ν_4	120	112	76	73	54	53
B_{1g}	ν_6	453	438	434	(291?)	430	(195?)
	ν_7	189	284	129	176	103	—
B_{2g}	ν_{11}	484	506	449	407	438	344
	ν_{12}	137	164	96	112	103	—
B_{3g}	ν_{15}	180	—	120	—	95	—

In each case the four most intense lines are assigned to the totally symmetrical class A_{1g}, and the twisting frequency ν_{15} is not observed, as was also found to be the case for B_2H_6. The only observed frequency not included in table 6 is the weak line of Al_2Br_6 at 221, which may be a combination tone $140 + 73 = 213$. (The corresponding combination tones for Al_2Cl_6 and Al_2I_6 would coincide almost exactly with the fundamentals 340 and 146 respectively.) It is also likely that the weak lines 291 (Al_2Br_6) and 195 (Al_2I_6) are not fundamentals, since they differ greatly from the calculated values, and can be represented as $176+112 = 288$ and $146+53 = 199$ respectively. The true fundamentals of ν_6 for Al_2Br_6 and Al_2I_6 may be obscured by ν_{11}.

The only unsatisfactory feature of the above scheme is that the lines assigned to ν_1 and ν_{11} for Al_2Cl_6 are described respectively as depolarized and polarized, while in theory the reverse should be the case (cf. table 1). The degree of polarization recorded for 506 is only 0·3, so that this objection is not a decisive one, but it is possible that the assignments of ν_1 and ν_{11} should be reversed.

Kohlrausch & Wagner (1942) have given assignments for the aluminium halides which differ considerably from ours, though they were arrived at by a similar treatment. This is because three of their four frequency equations contain errors, which led them to suppose that ν_{12} was not observed, and to interpret two of the strong observed lines as combination tones.* Kohlrausch & Wagner have also measured the Raman spectra of C_2Cl_6 and aluminium trimethyl. If the C_2Cl_6 spectrum is compared with those of the aluminium halides, the differences in general pattern are so marked as to exclude an ethane-like structure for the latter: this agrees with the results of electron-diffraction measurements (Palmer & Elliott 1938; Brode 1940). In the spectrum of aluminium trimethyl the higher frequencies can be attributed to internal motions of the CH_3 groups, while the low-frequency end of the spectrum differs considerably from that of C_2Cl_6 and shows some resemblance to the aluminium halide spectra. Kohlrausch & Wagner have concluded that dimeric aluminium trimethyl has the bridge structure  ${}^{\mathrm{Me}}_{\mathrm{Me}}\!>\!\mathrm{Al}\!<^{\mathrm{Me}}_{\mathrm{Me}}\!>\!\mathrm{Al}\!<^{\mathrm{Me}}_{\mathrm{Me}}$, like the dimeric halides. However, in this case, electron-diffraction measurements (Brockway & Davidson 1941) are definitely in favour of an ethane rather than a bridge structure. Moreover, the evidence for the existence of a definite species Al_2Me_6 is not strong, since although the measured vapour (Laubengayer & Gilliam 1941) corresponds roughly to dimerization at the lowest temperature used, the plot of apparent molecular weight against temperature has not flattened out, suggesting the existence of higher polymers. We have therefore not attempted to interpret the spectrum of this compound.

Our best thanks are due to Dr J. W. Linnett for his advice and assistance.

REFERENCES

Anderson, T. F. & Burg, A. B. 1938 *J. Chem. Phys.* **6**, 586.
Bauer, S. H. 1942 *Chem. Rev.* **31**, 43.
Bell, R. P. 1944 *Proc. Roy. Soc.* A, **183**, 328.
Bethe, H. 1933 *Handbuch d. Physik*, **24**, 434.
Brockway, L. O. & Davidson, N. R. 1941 *J. Amer. Chem. Soc.* **63**, 3287.
Brode, A. 1940 *Ann. Phys., Lpz.*, (5), **37**, 344.

---

* Two of the three incorrect equations were subsequently corrected (Wagner 1943), but no revision was made in the calculated frequencies or in the assignment proposed.

Crawford, B. L. & Edsall, J. T. 1939 *J. Chem. Phys.* **7**, 223.
Diatkina, M. E. & Sirkin, J. K. 1942 *C.R. Acad. Sci. U.R.S.S.* **35**, 180.
Douglas Clark, C. H. 1934 *Phil. Mag.* (7), **18**, 459.
Gerding, H. & Smit, E. 1941 *Z. Phys. Chem.* B, **52**, 171.
Gerhard, S. L. & Dennison, D. M. 1933 *Phys. Rev.* **43**, 197.
King, G. W. 1938 *J. Chem. Phys.* **6**, 378.
Kohlrausch, K. W. F. & Wagner, J. 1942 *Z. Phys. Chem.* B, **52**, 185.
Laubengayer, A. W. & Gilliam, W. F. 1941 *J. Amer. Chem. Soc.* **63**, 477.
Linnett, J. W. 1938 *J. Chem. Phys.* **6**, 692.
Longuet-Higgins, H. C. & Bell, R. P. 1943 *J. Chem. Soc.* p. 250.
Palmer, K. J. & Elliott, N. 1938 *J. Amer. Chem. Soc.* **60**, 1852.
Placzek, G. 1934 *Handbuch der Radiologie*, **6**, part 2.
Rosenbaum, E. J. 1940 *J. Chem. Phys.* **8**, 643.
Sirkin, J. K. & Diatkina, M. E. 1941 *Acta Physicochim. U.R.S.S.* **14**, 547.
Stitt, F. 1940 *J. Chem. Phys.* **8**, 981.
Stitt, F. 1941 *J. Chem. Phys.* **9**, 780.
Wagner, J. 1943 *Z. Phys. Chem.* B, **53**, 85.
Wiberg, E. & Johannsen, T. 1942 *Angew. Chem.* **55**, 38.
Wilson, T. P. 1943 *J. Chem. Phys.* **11**, 369.

The dissociation energies of CO, N_2, NO and CN

BY A. G. GAYDON AND W. G. PENNEY
Imperial College of Science and Technology, S.W. 7

(*Communicated by Sir Alfred Egerton, Sec.R.S.—Received* 30 *March* 1944)

No previous set of values of the dissociation energies of CO, N_2, NO and CN can be regarded as satisfactory unless violations of the non-crossing rule are postulated as possible.

The Birge-Sponer method of extrapolating vibrational energy levels to the dissociation limit is discussed. By assuming that the intensity variations in the Fourth Positive bands of CO, usually regarded as due to a predissociation, are really perturbations, and by adopting an alternative interpretation to the commonly accepted one for the predissociation in the First Positive bands of N_2, it is possible to reconcile existing data with both the non-crossing rule and the Birge-Sponer extrapolations only with the following values $D(\mathrm{CO}) = 11{\cdot}11$ eV, $D(\mathrm{N_2}) = 9{\cdot}764$, $D(\mathrm{NO}) = 6{\cdot}49$, $D(\mathrm{CN}) = 7{\cdot}5$. The corresponding value of the latent heat of sublimation of carbon is 170·1 kcal./mol. This scheme of values explains the results of related experimental measurements at least as well as any other.

INTRODUCTION

The energies of dissociation of a number of diatomic molecules have been determined, apparently with high accuracy, by observation of predissociation limits. In some cases, the reliability of the values obtained depends on the validity of certain assumptions concerning the degree of excitation of the products of dissociation, and in some cases the values at present accepted are very different from the values expected from extrapolation of known vibrational energy levels to the dissociation limit.

The recent history of the supposed determinations of the energy of dissociation of carbon monoxide, a molecule whose spectrum is well known and has been the subject of intensive study, illustrates how difficult it has proved to reach agreement. The value for D(CO) originally put forward by Birge & Sponer (1926) was, with recent conversion factors, 11·34 eV. Lessheim & Samuel (1934) assumed that there was a predissociation in the upper level of the 5B bands and gave D(CO) = 10·45 eV. Then Coster & Brons (1934) reported predissociation in the spectra of CO leading to D(CO) = 9·82 eV, but later (Brons 1935; Coster & Brons 1935) claimed to have observed predissociation in the $A^1\Pi$ upper level of the Fourth Positive bands and said that 'this made D(CO) = 8·41 eV perfectly certain'. This was at first accepted by Schmid & Gerö (1935) who later, however, appeared convinced that the true value was 6·92 eV (Schmid 1936; Schmid & Gerö 1936). Herzberg (1937) reviewed the existing observations of predissociations and concluded that, with recent conversion factors, values of 6·92, 7·88, 8·42, 9·14, 9·85 or 11·11 eV were possible according to the genuineness and interpretation of the observed limits, and summarized in favour of 9·14, a value which he has recently (1942) defended. Calculations on the short-wave fluorescence limits of benzene and of the absorption spectra of polyenes (Kynch & Penney 1941) favour a value about 11 eV, while electron collision experiments (Hagstrum & Tate 1941) gave 9·6, favouring a value of 9·85 if the predissociation limits are taken into account. From this remarkable collection of diverse values it does not seem possible to draw any safe conclusion, except perhaps that the whole position is unsatisfactory. For nitrogen a rather similar history could be written, but in this case a value of 7·38 eV appears to have been generally accepted, although it is in strong disagreement with the value 11 eV, obtained by extrapolation of the vibrational levels of the ground state, a value which can hardly be reduced by more than 2 eV on any normal scheme of extrapolation.

Quite apart from the obvious doubt engendered by the variety of values put forward, it is remarkable that most authors have ignored the non-crossing rule introduced by Hund, and stressed by Neumann & Wigner (1929) that potential energy curves of the same electronic species never cross. We therefore propose to apply the non-crossing rule systematically and see where it leads. As a preliminary to a discussion of individual molecules, a few brief general remarks will serve to express our views on the reliability of extrapolated dissociation limits and on the application of the rule. It is then proposed to discuss the dissociation energies of certain molecules affected by our conclusions. These energies are of considerable importance in thermo-chemistry, and the value of D(CO) carries with it, as described by Herzberg (1937), a knowledge of the heat of sublimation of carbon, and is therefore of cardinal importance in determining the absolute heats of formation of all organic compounds.

THE BIRGE-SPONER EXTRAPOLATION

When a number of vibrational energy levels of a molecular state are known, it is possible to extrapolate these levels to the convergence limit correlated with the energy of dissociation of that particular molecular state. The extrapolation may be made either analytically or graphically. One of the simplest methods is to plot the intervals $\Delta G_{v+\frac{1}{2}}$ between the vibrational energy levels against the vibrational quantum number v and extrapolate the curve to zero $\Delta G_{v+\frac{1}{2}}$. The dissociation energy is then given by the area beneath the curve. Alternatively the vibrational energy term G_v may be plotted against ΔG_v; the intercept on the G_v axis obtained by extrapolation then gives the dissociation limit. The accuracy of the extrapolation naturally depends on how many vibrational levels are known, and where only a few levels are known, so that a long extrapolation has to be made, the accuracy may not be high; but where a number of levels are known so that the general trend of the curve can be seen and allowed for, reasonably accurate values may be expected.

In a detailed study of the vibrational levels of the upper electronic state of O_2 Birge (1929) found that the $\Delta G_{v+\frac{1}{2}}-v$ curve showed an inflexion at high values of v, giving a slightly *higher* value for the dissociation limit than that obtained by extrapolation using only the lower levels. The second difference in $\Delta G_{v+\frac{1}{2}}$ showed a fairly sharp break, and Birge represented the curve for this second difference by two intersecting straight lines. More recent measurements by Knauss & Ballard, however, indicate that the change of slope is less sudden than was formerly supposed. Similar changes leading to slightly higher values of the dissociation energy have been observed for the upper electronic states of the halogens, ICl and IBr. The upper electronic states of the alkali-metal hydrides are also anomalous, due to interaction with ionic terms, but do not show any sharp discontinuity. Watson & Koontz (1934) have reported a break in the values for the second difference in $\Delta G_{v+\frac{1}{2}}$ for the upper state of the far ultra-violet system of N_2^+ which would in this case lead to a slightly lower value of the dissociation limit. However, all these discontinuities are slight and would not appreciably affect the determination of the limit.

For AgH (Bengtsson & Olsson 1931) the curve for the excited electronic state shows a pronounced change of slope which leads to a much higher dissociation limit than that which would be obtained by extrapolation from the first few levels.

The $\Delta G_{v+\frac{1}{2}}-v$ curve for the upper state of the red bands of K_2 studied by Loomis & Nusbaum (1932) shows a rather marked change of slope at high v, leading to a lower dissociation limit than that which would be obtained by extrapolation of the first part of the curve. Many of the weak bands from which data for the highest vibrational levels are derived are blended, while others could be analysed so as to form parts of weaker Franck-Condon parabolas of the type observed by Gaydon & Pearse (1939) for RbH. Thus it is hardly safe to rely on this example of a break in the curve towards a lower limit. There are, however, indications of similar changes in the curves for the excited states of other alkali-metals.

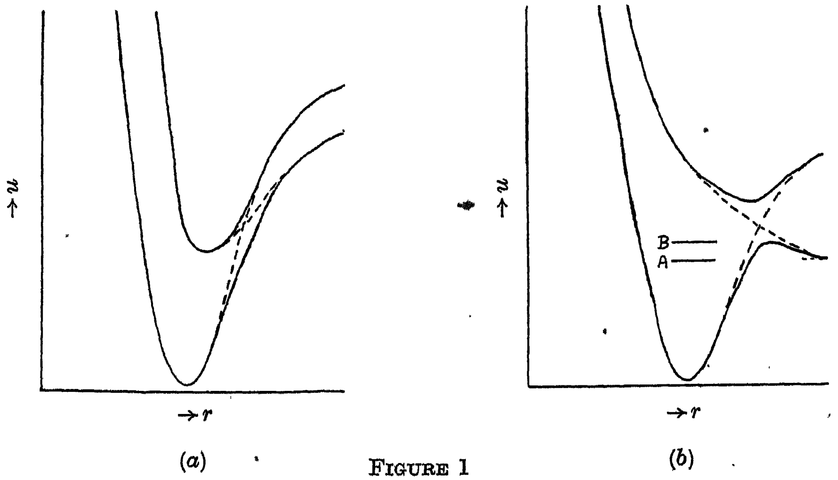
It should be noted that all the discontinuities discussed above refer to excited states. The authors are unaware of any evidence for a sharp change of slope in the curve for any *ground* electronic state, although a large number of such states have been examined. For some molecules, especially CO and N_2, many of the higher electronic states appear to possess only a few vibrational levels. In such cases, the vibrational levels seem to terminate abruptly, probably owing to predissociation caused by crossing of a repulsive potential curve, and they do not converge to a limit in the usual way.

We conclude that there is little evidence that the extrapolation to the dissociation limit, where it is made with reasonable care from a fair number of vibrational levels, is likely to be seriously in error for ground electronic states. The possibility of abrupt termination of vibrational levels due to crossing of the potential curve with a repulsive potential curve must however be borne in mind, especially for excited electronic states.

THE NON-CROSSING RULE

According to the non-crossing rule of Hund, and Neumann & Wigner (1929), potential energy curves of the same electronic species which would to a first approximation cross each other, will, in higher approximation, avoid this crossing.

Two cases of avoided crossing are illustrated in figure 1 (*a*) and (*b*). The broken lines indicate the potential curves in first approximation where they cross, while the full-line curves represent the higher approximation where the crossing is avoided by the repulsion between the two curves.



(*a*) FIGURE 1 (*b*)

It will be seen that where one of the curves, in first approximation, is repulsive, there is the possibility of the lower curve showing a maximum as well as the usual minimum characteristic of a stable state. We suspect that such maxima may be fairly common for molecules like CO and N_2 where we have evidence from the abrupt termination of vibrational levels that, at any rate among the higher electronic states, there is considerable interference between the numerous potential energy curves.

It should be noted that in an avoided crossing of the type illustrated in figure 1 (*a*) there is little chance of attributing vibrational levels of the upper curve to a continuation of the vibrational levels of the lower curve and so failing to distinguish between the molecule in the two states as, in general, the rotational constants and the internuclear distance will not vary smoothly in passing from one electronic state to the other; the value of B will usually be appreciably smaller for vibrational levels of the upper arm than for those of the lower. In the actual neighbourhood of an attempted crossing the concept of potential curves probably has little meaning and only the energy levels may be regarded as real, but away from the actual crossing and its accompanying perturbations we may expect the vibrational and rotational energy levels to behave normally with different constants for the two states.

In the case of a potential maximum, as illustrated in figure 1 (*b*), vibrational levels may exist above the dissociation limit, that is between the levels A and B in the figure. Molecules in these levels will therefore be in a metastable state and may be able to predissociate spontaneously. However, it does not always follow that their life will be short as, if the potential maximum is broad, there will be little overlap between the wave functions corresponding to the vibrational levels and the wave functions corresponding to the dissociated condition, and passage through the potential barrier may be very rare. Thus levels not much above A may behave essentially normally and strong transitions to and from these levels may be observed in the spectrum.

As far as we are aware there is only one example, CaH, in which there is actual data indicating crossing of potential energy curves of the same species, and this appears to be due to an incorrect interpretation of the spectrum rather than to a violation of the rule.* For CO, N_2, NO and CN, however, the use of low values for the dissociation energy leads to a violation of the rule because known potential curves would have to cross repulsive curves of the same species arising from the lowest dissociation limits.

In considering the non-crossing rule in relation to the determination of dissociation limits we may well quote the caution given by Mulliken (1937):

'...it would be wise to keep in mind the possibility of avoided crossing with fulfilment of the non-crossing rule and accompanied by maxima in the potential energy curves of the states under consideration. It is likely that maxima, small or large, will prove to be more common than has been supposed.

---

* For CaH, Grundstöm shows potential curves (reproduced by Herzberg 1939) in which that for the D $^2\Sigma^+$ state crosses that for C $^2\Sigma^+$. In Grundstöm's paper (1932) on the D system, several of the bands are found to show two extra branches (referred to as P^* and R^*) which cannot be accounted for on the existing analysis attributing the bands to a $^2\Sigma \rightarrow {}^2\Sigma$ transition. One of us (A. G. G.) has rephotographed the (1, 1) band on the 2nd order of a 20 ft. concave grating spectrograph. The existence of the P^* and R^* branches is confirmed. It is also noted that the lines denoted $P_1(17)$ and $R_1(15)$ by Grundstöm, are in reality double. This probably indicates that all the lines are really unresolved doublets, but only in one case, due to perturbation, is the doubling visible. The transition is therefore probably $^4\Sigma \rightarrow {}^2\Sigma$ and not $^2\Sigma \rightarrow {}^2\Sigma$. Thus the D level is not of the same species as the C level.

'If so, some of the existing conclusions as to dissociation energies based on predissociation phenomena or on convergence of vibrational series to a limit may need revision.

'...Even in cases such as O_2, N_2 and NO there remains a shadow of doubt as to the reliability or at least the accuracy of the accepted values. For instance, the existence of hitherto unsuspected maxima in potential curves might affect some of these values; also, the use of predissociation data appears to be somewhat risky unless great caution has been used.'

THE DISSOCIATION OF CO

The history of the determinations of $D(\text{CO})$ has been given briefly in the introduction. The $\text{B}\,^1\Sigma$ upper level of the Angström bands and the $b\,^3\Sigma$ upper level of the Third Positive bands both show definite predissociation at 11·11 eV, and this is certainly an upper limit for $D(\text{CO})$. Gerö (1936) reported predissociation in bands of the Fourth Positive system with $v' = 9$ and 10. Herzberg (1937) has discussed the three possibilities (*a*) that this predissociation is real and corresponds to a real limit, (*b*) that it is real but is accidental and does not correspond to a dissociation limit, and (*c*) that the predissociation is not real at all but is only a perturbation. The first possibility gives $D(\text{CO}) = 6{\cdot}92$ eV, a value advocated by Schmid and Gerö, the second possibility gives 9·14 eV, the value which Herzberg favours, while the third possibility, taken with the assumption that the predissociation at 11·11 is actually a limit, gives $D(\text{CO})$ as either 11·11 or 9·85 eV. Gerö's published photo showing this predissociation in the Fourth Positive bands, and our own spectrograms taken on a Hilger E.1 quartz Littrow spectrograph, seem to us to favour the third possibility, that the effect observed is only a perturbation and not a predissociation. Further, the superposition of a '3A' band on the (9, 18) Fourth Positive band (Gerö 1938) results in anomalous intensities of the lines of the latter, giving an effect resembling predissociation.

Herzberg (1942) has used Faltings, Groth and Harteck's observations on the photochemical decomposition of CO as an argument in favour of $D(\text{CO}) = 9{\cdot}14$. We have already discussed this photodissociation (Gaydon & Penney 1942) and pointed out that with due consideration for the initial act of absorption this does not rule out 9·85 eV, even if the initial act is direct photodissociation, but that probably it is a case of induced photodissociation and therefore tells us nothing. Electron collision experiments indicate $D(\text{CO}) = 9{\cdot}6 \pm 0{\cdot}2$ eV, but the curve from which this result is obtained appears to be of complex structure and might be capable of an alternative explanation. Certain chemical calculations involving the latent heat of sublimation of carbon, and also Marshall and Norton's direct determination of the heat of sublimation favour $D(\text{CO}) = 11$. Extrapolation of the vibrational levels of CO^+ favour a value a little over 10 eV (but see footnote added in proof*).

We now propose to examine the application of the non-crossing rule to the various potential energy curves for CO. The lowest possible dissociation products are $C(^3P)+O(^3P)$. These lead, on the Wigner-Witmer correlation rules, to the formation of molecular states of the following types

$$^1\Sigma^+\,(2),\quad ^1\Sigma^-,\quad ^1\Pi\,(2),\quad ^1\Delta,\quad ^3\Sigma^+\,(2),\quad ^3\Sigma^-,\quad ^3\Pi\,(2),\quad ^3\Delta$$

together with a number of quintet states. Now, if the non-crossing rule is to be upheld, these molecular states arising from the lowest products of dissociation must either link up with the various known electronic levels of CO of corresponding type or must lie wholly below them so as to avoid crossing. An energy level diagram is shown in figure 2; this gives all the known electronic states of CO with the certainly known vibrational energy levels indicated for each. For application of the non-crossing rule, the $a'\,^3\Sigma^+$ state is the most interesting. From the analysis of the Asundi bands, some 12 vibrational levels are known, but by following up perturbations in the $A\,^1\Pi$ upper state of the Fourth Positive bands and in the $B\,^1\Sigma^+$ upper state of the Ångström bands, Schmid & Gerö (1937) have been able to track the vibrational levels up to $v'=41$ for $a'\,^3\Sigma^+$ (the v' values used in the original paper need to be raised by 2 units). The vibrational levels are smoothly spaced up to this point which is within about 1000 cm.$^{-1}$ of the extrapolated convergence limit at 11·1 eV. Now the normal dissociation products must result in two $^3\Sigma^+$ states, i.e. two states of the same species as $a'\,^3\Sigma^+$. The regularity of the vibrational levels of this state does not indicate any perturbations such as would result from an avoided crossing, so that if the rule is to hold $a'\,^3\Sigma^+$ must dissociate to normal products, or alternatively the two potential curves of $^3\Sigma^+$ type from these products must lie below that for $a'\,^3\Sigma^+$. In the latter case there must be two unknown $^3\Sigma^+$ states lying below 50,000 cm.$^{-1}$, and this is inconceivable. It would be just possible that the curve for $a'\,^3\Sigma^+$ might reach a maximum at 11·1 eV, and then descend to a limit below this point, but the exact coincidence of this extrapolated limit at 11·1 with the energy of the two known strong predissociations is too striking for this hypothesis to appear tenable. If the non-crossing rule holds there can be no doubt that $D(CO) = 11\cdot1$ eV.

It may be noted that with the value for $D(CO)$ favoured by Herzberg (9·14 eV) the $A\,^1\Pi$ state also violates the non-crossing rule.

---

* (*Footnote added in proof.*) A recent letter by Anand (1942), reporting Rydberg series in the far ultra-violet spectrum of CO, suggests that $I(CO) = 14\cdot55$ eV; this would give $D(CO^+) = D(CO) - 3\cdot33$. The discrepancy between the extrapolated values for the various states of CO^+ and those to be expected using $D(CO) = 11\cdot11$ would thus be reduced appreciably.

It may be noted that while a neutral molecule dissociates to uncharged atoms, an ionized molecule such as CO^+ or N_2^+ dissociates to one charged and one uncharged atom. It would be expected that the law of force between a charged and an uncharged atom would, at large internuclear distances, differ from that between two uncharged atoms. Thus it is not unlikely that for high values of the vibrational quantum number the form of the Birge-Sponer extrapolation for an ionized molecule may differ from that of an ordinary neutral molecule.

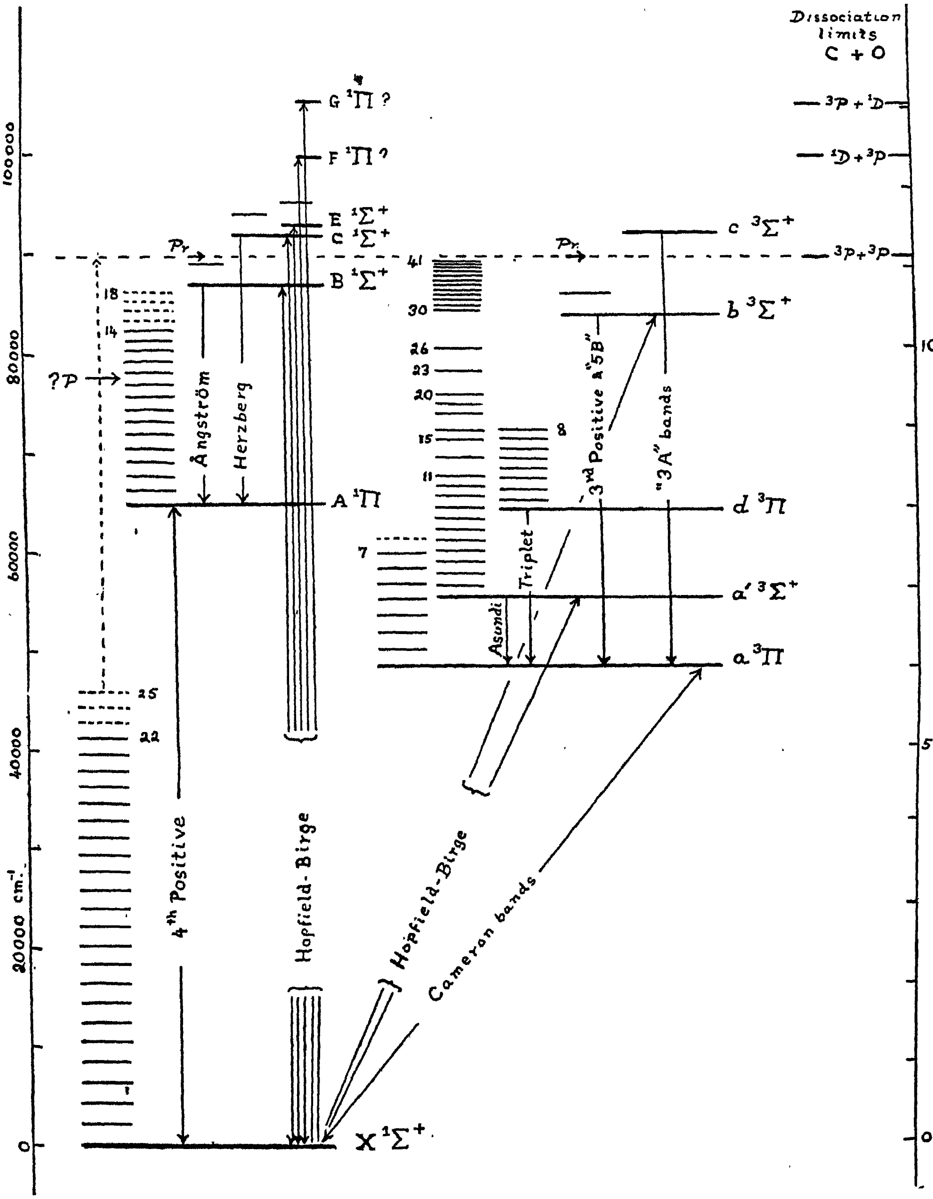


FIGURE 2. Energy level diagram of CO. The main predissociations are indicated by *Pr.* The intensity fluctuations produced by irregularities in the A $^1\Pi$ state, which we believe to be a perturbation and not a predissociation, are indicated by *?P*. The data for $d\,^3\Pi$ are from Gerö & Szabó (1939); their analysis of the Triplet system does not, in our opinion, lead to a satisfactory intensity distribution; the *v* enumeration for $d\,^3\Pi$ probably requires adjustment.

THE DISSOCIATION OF CO^+

The value of $D(CO^+)$ can be determined from the cycle

$$D_0(CO^+) = D_0(CO) + I(C) - I(CO).$$

Taking the ionization potential of carbon as 11·22 and of carbon monoxide as 14·1 (Hagstrum & Tate 1941) we get $D(CO^+) = D(CO) - 2{\cdot}9$.* In view of the doubt about CO it seems desirable to see if we can obtain any additional evidence from a study of the spectrum of CO^+.

Three band systems are known for CO^+, the First Negative carbon bands B $^2\Sigma$ to X $^2\Sigma$, the Comet-tail system A $^2\Pi$ to X $^2\Sigma$, and the Baldet-Johnson system B $^2\Sigma$ to A $^2\Pi$. From the analysis of these systems some thirteen vibrational levels, the highest with an energy of 25971 cm.$^{-1}$ (= 3·21 eV), are known for the lowest, X $^2\Sigma$, electronic state. These levels show an almost constant second difference, and Biskamp (1933) has made a linear extrapolation to a limit at 79700 cm.$^{-1}$ = 9·9 eV, a value which, in view of the length of the extrapolation may not be of high accuracy. For the upper B $^2\Sigma$ state, which lies at an energy of 45652 cm.$^{-1}$ (5·66 eV) above the ground state, ten vibrational levels with a total energy of 14543 cm.$^{-1}$ (1·80 eV) are known, these extrapolating, according to Biskamp, to 29600 cm.$^{-1}$, that is about 75252 cm.$^{-1}$ or 9·3 eV above the ground state. The A $^2\Pi$ state lies some 15600 cm.$^{-1}$ (= 1·93 eV) above the ground state. Fourteen vibrational energy levels are known extending to 45762 cm.$^{-1}$ (= 5·67 eV) above the ground state. We have extrapolated these vibrational levels to their dissociation limit, and assuming a normal convergence find a value of about 9·2 eV for this limit.

The lowest dissociation products for CO^+ are $C^+(^2P)$ and $O(^3P)$, and this combination will produce one $^2\Sigma^+$, two $^2\Sigma^-$, two $^2\Pi$ and other molecular states of higher multiplicity and Λ value. There does not appear to be any experimental evidence as to whether the observed $^2\Sigma$ states of CO are + or −. If they were − then there would be no difficulty in assuming that all the three known states converged together to normal products at a dissociation limit of around 9 eV giving a high value for $D(CO)$. However, it is usually assumed that both $^2\Sigma$ states are $^2\Sigma^+$, an assumption which seems more reasonable *a priori*. If this is so then X $^2\Sigma$ and B $^2\Sigma$ must dissociate to different limits. To avoid crossing, the X $^2\Sigma$ must dissociate to $C^+(^2P) + O(^3P)$ while B $^2\Sigma$ must go to $C^+(^2P) + O(^1D)$ some 1·97 eV higher. Taking the limit for B $^2\Sigma$ as 9·3 this would give $D(CO^+)$ as 7·3 and $D(CO)$ as 10·2; to bring this limit up to that required to give $D(CO) = 11{\cdot}1$ it would be necessary to postulate a fairly considerable break in the $\Delta G_v - v$ curve for B $^2\Sigma$, as is in fact observed for one of the states of N_2^+. On the other hand the vibrational levels of A $^2\Pi$ which should, on the non-crossing rule, dissociate to normal products, are known up to 5·67 eV and appear as shown above to converge to a value of $D(CO^+)$ a good deal higher than 7·3. It is probable that for CO^+ the potential curves for the two known excited electronic states are perturbed by other states and depart from the conventional form. All that may fairly be deduced is that data obtained from CO^+ favour a value for $D(CO)$ above 10 eV.

---

* See footnote on p. 380.

THE DISSOCIATION ENERGY OF N_2

The spectrum of nitrogen shows definite predissociation in bands of the Second Positive system with $v' = 2$, 3 and 4. This leads to a dissociation limit at 12·145 eV. Van der Ziel (1934, 1937) has also found a weak predissociation in some of the upper vibrational levels of the $B\,{}^3\Pi_g$ state, which is the initial state of the First Positive system. This predissociation occurs at 9·844 eV. Both of the components of the Π levels resulting from Λ doubling are affected by the predissociation, and the strength of the predissociation increases with rotational quantum number. This has led van der Ziel to interpret the predissociation as due to a repulsive Δ state, and this argument leads to $D(N_2) = 7{\cdot}383$ eV.

As with CO, this low value of the dissociation energy implies a violation of the non-crossing rule. Some sixteen vibrational levels are known for the $A\,{}^3\Sigma_u^+$ state, these rising to about 1 eV above the dissociation limit $N({}^4S)+N({}^4S)$ which should give rise to a ${}^3\Sigma_u^+$ state. Similarly, vibrational energy levels for $a\,{}^1\Pi_u$ and $B\,{}^3\Pi_g$ are known well above the $N({}^4S)+N({}^2D)$ limit from which states of these types should arise. An energy level diagram for N_2 showing this is given by Herzberg (1939), who has accepted van der Ziel's interpretation of the predissociation.

This violation of the non-crossing rule led us (1942) to doubt the low value for the dissociation energy. It has now been shown by one of us (Gaydon 1944*a*) that there is an alternative interpretation of van der Ziel's predissociation. It has apparently been generally assumed that when a Π state is perturbed or predissociated by a Σ state only one of the two components of the Π state due to Λ doubling is affected. Actually this is invariably true only for Π states in Hund's case *b*. In Hund's case *a* the rotational quantum number J is the determining factor in the rotational energy of the Π state and in the selection rules, and K is not rigorously a good quantum number. The $B\,{}^3\Pi_g$ state of N_2 is intermediate between Hund's cases *a* and *b* but approximates sufficiently to the former to lead us to expect that the quantum number J will dominate K in the choice of selection rules.

The lowest dissociation products for N_2 are ${}^4S+{}^4S$, and from these atomic states a molecular ${}^5\Sigma_g^+$ state can arise. It is reasonable to suppose that this state is repulsive. Each rotational level (in so far as it is possible to define a rotation of a non-stable molecule) is defined by a quantum number K, and each level K consists of five spin components $J = K-2$, $K-1$, K, $K+1$ and $K+2$. The + or − symmetry of all five spin components of any K level is the same (Symmetry + or − refers to reflexion in the origin of the nuclear and electronic coordinates, inclusive of spin). For a proof of the statement see Wigner & Witmer (1928)).

In order to interpret van der Ziel's predissociation, we suppose that near the crossing of the potential curves of $B\,{}^3\Sigma_g$ and ${}^5\Sigma_g^+$ it is possible for a range of J values in the stable Π state to change spontaneously to the same J value in the Σ state. Any small adjustment in the rotational energy is easily accomplished by variations in the kinetic energy of the nuclei along the inter-nuclear direction, since this motion is not quantized in an unstable state. Each J value in the $^5\Sigma_g^+$ state occurs twice with one symmetry (+ or −) and thrice with the opposite symmetry (− or +). Hence both components of the Λ doublets of the Π state can predissociate.

It is perhaps worth pointing out that if the Σ state had been stable, the mutual perturbations of the Σ and Π states would be very much more limited because the condition of equal energy of levels with the same symmetry and J value will be satisfied only at one point in each series of levels. However, for CN, Wager (1943) has observed perturbations in both Λ components of a $^2\Pi$ state, in Hund's case a, resulting from interaction with a $^2\Sigma$ state. Levels of the $\mathrm{A}\,^2\Pi$ state, with vibrational quantum number $v = 9$, are perturbed by levels of the $\mathrm{B}\,^2\Sigma$ state with $v = 0$. The levels of $^2\Pi_{\frac{1}{2}}$ with $J = 3\frac{1}{2}$ and $7\frac{1}{2}$ and of $^2\Pi_{\frac{3}{2}}$ with $J = 10\frac{1}{2}$ and $15\frac{1}{2}$ interact with levels of the $^2\Sigma$ state with $K = 4, 7, 11$ and 15 respectively. Detailed examination of this series of perturbations shows that both components of the Π state (due to Λ doubling) are affected, but at different energies.

With this interpretation of van der Ziel's predissociation in N_2, and making use of the definite dissociation limit at 12·145 eV, we obtain for the dissociation energy a value of 9·764 eV. The values of 7·383 and 8·573 are not ruled out by this interpretation of van der Ziel's predissociation, but, as already stated, the 7·383 value leads to a violation of the non-crossing rule by the $\mathrm{B}\,^3\Sigma_g$ and $a\,^1\Pi_u$ states of N_2, while the 8·573 value implies a violation of the rule for several states of NO (see later). No violation of the non-crossing rule is indicated by the value of 9·764, which also agrees better with the values obtained by the Birge-Sponer extrapolation for the $\mathrm{X}\,^1\Sigma^+$, $a\,^1\Pi_u$, $\mathrm{A}\,^3\Sigma_u^+$ and $\mathrm{B}\,^3\Pi_g$ states of N_2.

It should be noted that Hagstrum & Tate (1941) have made a determination of the dissociation energy by electron collision experiments using a mass spectrometer. They found that ionized nitrogen atoms appeared at 24·3 eV. Taking the ionization potential of atomic nitrogen as 14·55 and using the relation

$$A(\mathrm{N}^+) = I(\mathrm{N}) + D(\mathrm{N}_2)$$

we obtain directly $D(\mathrm{N}_2) = 9{\cdot}75 \pm 0{\cdot}2$ eV, in good agreement with the above value of 9·764 eV. In order to bring their results into line with the spectroscopic value of 7·383, Hagstrum & Tate assumed that the products of dissociation were excited atoms, but this now seems unnecessary.

THE DISSOCIATION ENERGY OF N_2^+

The dissociation energy of ionized nitrogen is linked with that of the neutral molecule by the cycle,

$$D(\mathrm{N}_2) + I(\mathrm{N}) = I(\mathrm{N}_2) + D(\mathrm{N}_2^+).$$

Using the latest values for $I(\mathrm{N})$ and $I(\mathrm{N}_2)$ as 14·55 and 15·58 we obtain

$$D(\mathrm{N}_2) - D(\mathrm{N}_2^+) = 1{\cdot}03.$$

The First Negative bands of N_2^+ correspond to a $^2\Sigma_u^+ \rightarrow {}^2\Sigma_g^+$ transition, and extrapolation of the vibrational levels indicates that both states probably have the same dissociation limit which is presumably the lowest, $N(^4S)+N^+(^3P)$. Extrapolation of the vibrational levels of the lower state is of little value as the extrapolation to the limit is long. For the upper level, the situation is more favourable. Herzberg (1928) observed definite vibrational levels up to $v' = 11$ and using data for *band heads*, obtained $D(N_2^+)$ around 6·9 eV. Since then Parker (1933) has made a rotational analysis of the (13, 15) First Negative band. Making use of this additional data, and using values (calculated with the aid of the rotational constants) for *band origins* instead of heads (which differ significantly because some bands are degraded to the red and some to the violet and in some cases the heads and origins are far apart) we have repeated the extrapolation. The result is shown in figure 3, which is a curve between $\Delta G_{v+\frac{1}{2}}$ and the total vibrational energy term G_v. The values of $D(N_2)$ to which the extrapolation would lead are indicated along the G_v axis. It will be seen that the curve, like that of AgH, shows a definite inflexion leading to a higher dissociation energy and that while a value for $D(N_2)$ as high as 9·7 would have appeared unlikely from extrapolation of the lower vibrational energy levels, with the new data for higher levels, this value does appear possible.

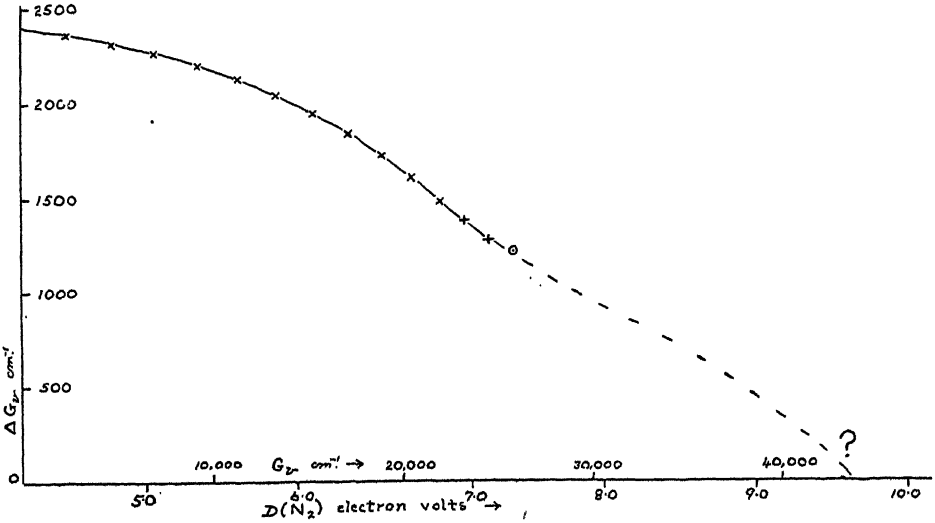


FIGURE 3. Extrapolation of the vibrational levels of the upper, $^2\Sigma^+$, state of the N_2^+ First Negative bands to the dissociation limit.

× × × Points obtained from Herzberg's measurements (origins).

+ + Points based on Parker's value for the origin of the (13, 15) band; this leads to a reliable value for G_{13}, and, since G_{11} is known, to the sum $\Delta G_{11\frac{1}{2}} + \Delta G_{12\frac{1}{2}}$; either of these points may be raised provided the other is lowered by an equal amount.

⊙ This point is based on the observation by Herzberg of the (14, 10) band at 3217·7 A; this is a weak band and this point may not be reliable.

It will be noted that a smooth extrapolation to $D(N_2) = 7{\cdot}38$ would be quite impossible, and indeed the $v' = 14$ level itself lies slightly above this energy.

We thus see that evidence obtained from the spectrum of ionized nitrogen does not contradict the new higher value for the dissociation energy of N_2. The value for N_2^+ is probably $9{\cdot}764 - 1{\cdot}03 = 8{\cdot}76$ eV.

THE DISSOCIATION ENERGY OF NITRIC OXIDE

The dissociation energy of nitric oxide is linked with that for nitrogen and oxygen by thermochemical data (see Bichowsky & Rossini 1936)

$$2NO = N_2 + O_2 + 1{\cdot}87 \text{ eV}.$$

Taking $D(O_2) = 5{\cdot}082$ and $D(N_2) = 9{\cdot}764$ this leads to $D(NO) = 6{\cdot}49$ compared with the lower value of 5·29 given by Herzberg (1939).

One of us (Gaydon 1944*b*) has recently studied the several band systems emitted by NO and has drawn up potential energy curves for all the known electronic states. It has been shown that it was extremely difficult to reconcile a low value of $D(NO)$ with the non-crossing rule. The new value lies just above all the known energy levels for the β, γ and δ systems, and although it is necessary to assume that the potential energy curves of the upper electronic states are largely determined by higher dissociation products, the non-crossing rule is not actually violated, and no energy levels are known above the expected dissociation limits for the respective states.

THE DISSOCIATION ENERGY OF CYANOGEN

Using certain thermochemical data, including a somewhat low value for the dissociation of C_2N_2 into 2CN, Herzberg (1939) obtained $D(CN) = 5{\cdot}96$ eV. Schmid, Gerö & Zemplén (1938) have followed vibrational levels for the $A\,^2\Pi$ state of CN up to $v = 30$ at an energy of about 51,000 cm.$^{-1}$ (6·3 eV) by observing perturbations in the violet system. The extrapolated limit for $A\,^2\Pi$ obtained by Schmid, Gerö & Zemplén is $60{,}500 \pm 1{,}000$ cm.$^{-1}$ ($7{\cdot}5 \pm 0{\cdot}12$ eV). Now on the non-crossing rule $A\,^2\Pi$ should dissociate to the lowest possible products, $C(^3P) + N(^4S)$, so that if $D(CN) = 5{\cdot}96$ we are again faced with a violation of the non-crossing rule.

However, if we use the new values $D(CO) = 11{\cdot}11$, $D(N_2) = 9{\cdot}76$ and take (see data summarized by Bichowsky & Rossini 1936)

$$C_2N_2 + 2O_2 = 2CO_2 + N_2 + 260 \text{ kcal.}$$

$$2CO + O_2 = 2CO_2 + 134 \text{ kcal.}$$

and use White's (1940) spectroscopically obtained value for the dissociation of C_2N_2 into 2CN as 146 ± 4 kcal., we obtain

$$D(CN) = 7{\cdot}6 \pm 0{\cdot}2 \text{ eV},$$

which is in satisfactory agreement with the direct spectroscopic limit obtained by Schmid, Gerö & Zemplén.

CONCLUSION

The only molecules which have hitherto been thought to violate the non-crossing rule (excepting the doubtful case of CaH) are CO, CN, NO and N_2, which are just those molecules for which the Birge-Sponer extrapolation was in serious disagreement with the dissociation energies obtained by predissociation data. It has been shown that by interpreting the effect in the Fourth Positive bands of CO as a perturbation instead of as a predissociation and by using the alternative explanation for van der Ziel's predissociation in the nitrogen spectrum, violation of the non-crossing rule can be satisfactorily avoided for all the above molecules.

We are therefore of the opinion that the non-crossing rule is a reliable guide in the correlation of atomic and molecular energy levels. The most probable values for the dissociation energies appear to be

$$\begin{aligned} D(\mathrm{CO}) &= 11{\cdot}11\ \mathrm{eV} = 256{\cdot}1\ \mathrm{kcal.},\\ D(\mathrm{N_2}) &= 9{\cdot}764\ \mathrm{eV} = 225{\cdot}1\ \mathrm{kcal.},\\ D(\mathrm{NO}) &= 6{\cdot}49\ \mathrm{eV} = 149{\cdot}5\ \mathrm{kcal.},\\ D(\mathrm{CN}) &= 7{\cdot}5\ \mathrm{eV} = 174\ \mathrm{kcal.},\\ \mathrm{C_2N_2} &= 2\mathrm{CN} - 6{\cdot}3\ \mathrm{eV}\quad (= 146\ \mathrm{kcal.}). \end{aligned}$$

Latest heat of sublimation of carbon 170·1 kcal.

One of us (A. G. G.) wishes to acknowledge generous financial assistance from the Warren Research Fund of the Royal Society.

REFERENCES

Anand, B. M. 1942 *Science and Culture*, **8**, 278.
Bengtsson, E. & Olsson, E. 1931 *Z. Phys.* **72**, 163.
Bichowsky, F. R. & Rossini, F. D. 1936 *The thermochemistry of the chemical substances.* New York: Reinhold.
Birge, R. T. 1929 *Trans. Faraday Soc.* **25**, 707.
Birge, R. T. & Sponer, H. 1926 *Phys. Rev.* **28**, 259.
Biskamp, H. 1933 *Z. Phys.* **86**, 33.
Brons, F. 1935 *Physica, Eindhoven*, **2**, 1108.
Coster, D. & Brons, F. 1934 *Physica, Eindhoven*, **1**, 155.
Coster, D. & Brons, F. 1935 *Amst. Proc. K. Akad.* **38**, 961.
Gaydon, A. G. 1944*a* *Nature*, **153**, 407.
Gaydon, A. G. 1944*b* *Proc. Phys. Soc.* **56**, 160.
Gaydon, A. G. & Pearse, R. W. B. 1939 *Proc. Roy. Soc.* A, **173**, 37.
Gaydon, A. G. & Penney, W. G. 1942 *Nature*, **150**, 406.
Gerö, L. 1936 *Z. Phys.* **99**, 52; **100**, 374.
Gerö, L. 1938 *Z. Phys.* **109**, 210.
Gerö, L. & Szabó, F. 1939 *Ann. Phys., Lpz.*, **35**, 597.
Grundström, B. 1932 *Z. Phys.* **75**, 302.
Hagstrum, H. D. & Tate, J. T. 1941 *Phys. Rev.* **59**, 354.

Herzberg, G. 1928 *Ann. Phys., Lpz.*, **86**, 189.
Herzberg, G. 1937 *Chem. Rev.* **20**, 145.
Herzberg, G. 1939 *Molecular spectra and molecular structure.* New York: Prentice-Hall.
Herzberg, G. 1942 *J. Chem. Phys.* **10**, 306.
Kynch, G. J. & Penney, W. G. 1941 *Proc. Roy. Soc.* A, **179**, 214.
Lessheim, H. & Samuel, R. 1934 *Proc. Phys. Soc.* **46**, 523.
Loomis, F. W. & Nusbaum, R. E. 1932 *Phys. Rev.* **38**, 1447.
Mulliken, R. S. 1937 *J. Phys. Chem.* **41**, 5.
v. Neumann, J. & Wigner, E. 1929 *Phys. Z.* **30**, 467.
Parker, A. E. 1933 *Phys. Rev.* **44**, 914.
Schmid, R. 1936 *Z. Phys.* **99**, 274.
Schmid, R. & Gerö, L. 1935 *Z. Phys.* **96**, 546.
Schmid, R. & Gerö, L. 1936 *Z. Phys.* **99**, 281.
Schmid, R. & Gerö, L. 1937 *Z. Phys.* **105**, 36.
Schmid, R., Gerö, L. & Zemplén, J. 1938 *Proc. Phys. Soc.* **50**, 283.
Wager, A. T. 1943 *Phys. Rev.* **64**, 18.
Watson, W. W. & Koontz, P. G. 1934 *Phys. Rev.* **46**, 32.
White, J. U. 1940 *J. Chem. Phys.* **8**, 459.
Wigner, E. & Witmer, E. E. 1928 *Z. Phys.* **51**, 859.
Ziel, A. van der 1934 *Physica, Eindhoven*, **1**, 353.
Ziel, A. van der 1937 *Physica, Eindhoven*, **4**, 373.

The structure of polyisoprenes

I. The crystal structure of geranylamine hydrochloride

BY G. A. JEFFREY

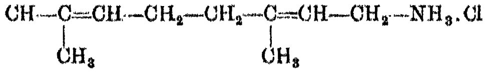
(*Communicated by W. N. Haworth, F.R.S.—Received* 25 *May* 1944)

An X-ray analysis employing three-dimensional Fourier syntheses has established the crystal structure and molecular dimensions of the di-isoprene derivative, geranylamine hydrochloride. The molecules, which have a *trans* configuration and are therefore analogous to gutta-percha, lie parallel and end to end in pairs within an ionic framework where each nitrogen atom is equidistant from four chlorine neighbours. The two isoprene units are planar and have normal interatomic distances, but are linked by a C-C bond markedly shorter than a normal single bond. This unusual bond feature is accompanied by a coplanar arrangement with the adjacent carbon bonds.

INTRODUCTION

The difficulty of crystal structure determination of complex organic molecules is to some degree offset by the extent of the diffraction data. With the polyisoprenes much of this information is obscured both by the diffuse nature of the fibre diagrams and by ambiguities arising from imperfect crystallite orientation. The resulting lack of precision is well illustrated by the example of rubber, where attempts at direct analysis have led to considerable disagreement even in the preliminary description of the unit cell (Lotmar & Meyer 1936; Sauter 1937; Morss 1938; Misch & Van der Wyk 1940). The detailed structures for rubber and β-gutta-percha recently proposed by Bunn (1942) are subject to the same limitations and cannot be regarded as necessarily final solutions, particularly as they entail surprisingly large valency distortions of a type hitherto unparalleled.

With the object of determining indirectly the interatomic dimensions in these polymers attention has been directed to simpler analogues. Following the study of a mono-isoprene compound by Cox & Jeffrey (1942), derivatives of geraniol were considered since they possess a di-isoprene system akin to the repeating units in the long-chain polymers. Although many of these compounds are unsuitable for X-ray methods, the hydrochloride of geranylamine (1) afforded the requisite crystals for intensity measurements, and their favourable space group together with the presence of a comparatively heavy chlorine atom has made possible the complete and accurate analysis of the structure:



$$\underset{\displaystyle CH_3}{\overset{}{CH-\underset{|}{C}}}\!=\!CH-CH_2-CH_2-\underset{\substack{|\\ \displaystyle CH_3}}{C}\!=\!CH-CH_2-NH_3.Cl \qquad (1)$$

EXPERIMENTAL

Four molecules of $C_{10}H_{18}NH_2.HCl$ are associated by the space-group symmetry $P2_1/c$ in a unit cell of dimensions $a = 22{\cdot}68$ A, $b = 5{\cdot}94$ A, $c = 8{\cdot}98$ A, $\beta = 98{\cdot}8°$. A complete description of the structure excluding the hydrogen atoms therefore requires the determination of 36 atomic parameters. For this purpose, all plane reflexions within the range of Cu K_α radiation were recorded on oscillation photographs about the principal axes, and their relative intensities were estimated by eye with the aid of calibration charts. As it is important that these values should be absolute, they were related to selected intensities transmitted through a thin crystal slab and measured by photometric comparison with a standard rock-salt (400) reflexion recorded on the same film. All photographs were taken three or more times with different exposures, using aluminium screens of known absorption to bring strong spots within the range of linear X-ray density.

From the experimental measurements the structure amplitudes of 1060 planes were calculated by the usual formulae, and were used first in the summation of a three-dimensional F^2 (Patterson) Fourier synthesis.

THE ARRANGEMENT OF THE CHLORINE ATOMS

The positions of the chlorine atoms were derived directly from the Patterson synthesis by virtue of the outstanding Cl-Cl vector peaks, two at co-ordinates (0·107, 0·50, 0·162) and (0·107, 0·50, 0·662) and a third of double height at (0·00, 0·00, 0·50). These peaks must arise from atoms at either (0·053, 0·25, 0·081) or (0·053, 0·25, 0·331) in conjunction with atoms symmetrically related. The former was ultimately shown to be correct, but both possibilities had to be considered since no distinction can be made between the centre of symmetry and the screw axis for atoms on the glide planes.

The Patterson synthesis, having symmetry planes at $y = 0$ and $\frac{1}{2}$, cannot reveal any small departure in the y parameter from 0·250. This was assumed to be exact for the approximate structure, and during the successive refinements no evidence to the contrary was observed. Further confirmation was available at the conclusion of the analysis from the comparison of the agreement between observed and calculated structure amplitudes for planes with l odd, where the chlorines make zero contribution, and those with l even:

$$\frac{\Sigma \,|F_{\text{obs.}}| - |F_{\text{calc.}}|}{\Sigma \,|F_{\text{obs.}}|} \quad 17{\cdot}0 \text{ and } 17{\cdot}8\,\% \text{ respectively.}$$

As shown in figures 4 and 5 these halogen atoms lie in sheets close to the (100) planes enclosing the organic molecules in pairs between them.

DETERMINATION OF THE APPROXIMATE STRUCTURE

The general orientation of the molecules was apparent from the length of the a axis; and from the weak odd layer lines on the c axis photographs it was clear that other atoms besides the halogens must lie on or near the glide planes. Further information was derived from the Patterson synthesis which was computed over sections parallel to (010) at intervals of y sufficiently close to reveal all prominent peaks. Six such peaks of height corresponding to Cl-N and Cl-C distances indicated positions for N, C_1, and C_2 in which the nitrogen was equidistant from four chlorine atoms and the interatomic distances and valency angles were approximately normal. This was the only arrangement compatible with the vector maps. Beyond this stage, however, these became much more complicated and further direct interpretation was not possible. For the remaining eight atoms trial and error methods were adopted and the Patterson synthesis was useful only in restricting the number of possibilities to be considered.

A model based on normal interatomic dimensions was used, first to ascertain arrangements which would satisfy the principal structure factors of the ($h0l$) zone, and then for further trials with general planes. Owing to the wide range of free rotation about each single bond, a large number of configurations was tried and rejected before arriving at a set of co-ordinates which, when refined by Fourier F synthesis, resulted in a marked improvement in the agreement between the observed and calculated structure factors. These and all subsequent syntheses were of the three-dimensional type, since projections were of little value even on the most favourable (010) plane.

THE FOURIER F SYNTHESIS

Starting from the approximate structure, the atomic co-ordinates were refined by the method of successive Fourier syntheses. The summations were evaluated from all the available data using Beevers and Lipson's strips over sections and along lines at values of x, y and z appropriate to the maximum of each atom peak. After

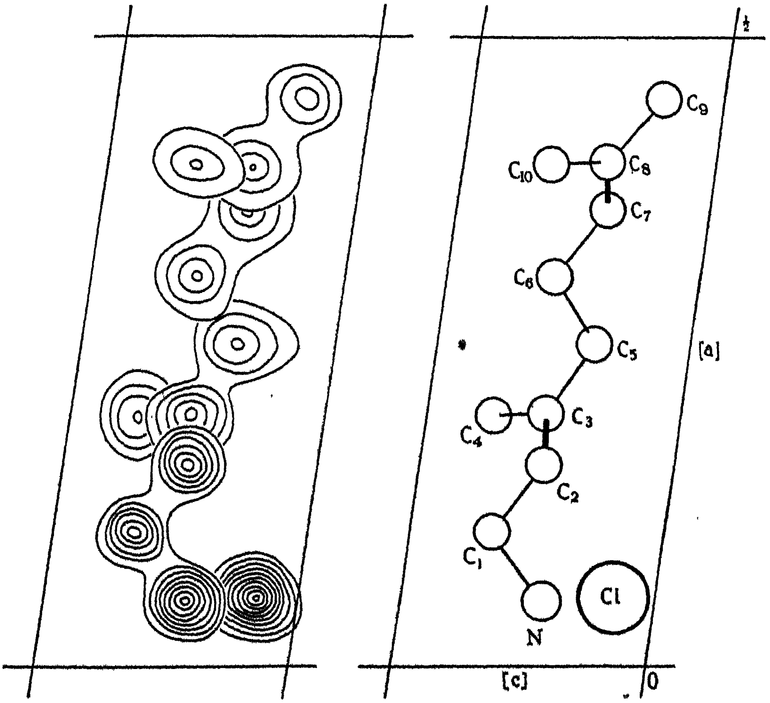


FIGURE 1. Projection of the molecule and Fourier sections on (010).

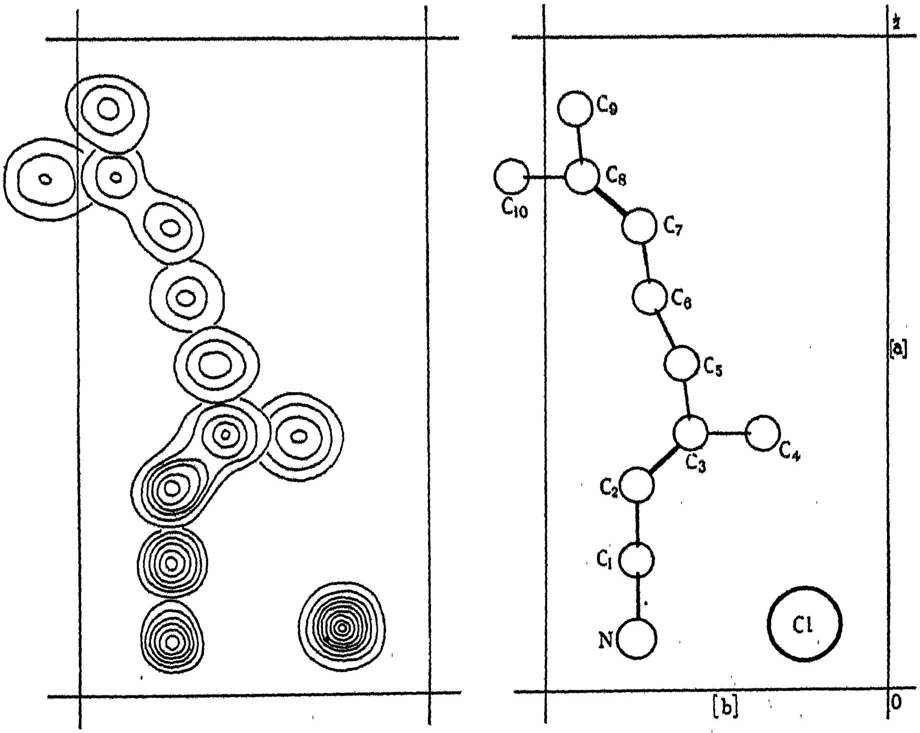


FIGURE 2. Projection of the molecule and Fourier sections on (001).

the fifth series, the structure-factor calculations revealed no further sign changes, and the electron maps from this final summation are shown in figures 1 and 2. For the purpose of illustration, the electron density peaks for one molecule have been taken from the sectional summations and superimposed on the *b* and *c* planes respectively. The contours are at intervals of one electron per A^3, with the exception of the chlorine atom where the scale is two electrons. The zero and first contours have been omitted for clarity in the diagrams. These maps show clearly that the peaks are different for carbon atoms at opposite ends of the molecule and although approximately equal in total volume they become progressively flatter for each atom along the chain from the nitrogen. While it cannot be certain that this has a real physical significance and does not arise merely from the use of a finite Fourier series, the magnitude of the effect suggests that it may be attributed to a variation in thermal motion between the polar and non-polar ends of the molecule.

THE FINAL ATOMIC PARAMETERS

The atomic parameters were measured directly from the maximum of each peak in the Fourier synthesis and gave the interatomic dimensions shown in table 1.

TABLE 1

N-C_1	1·49 A	(1·48)	C_5-C_6	1·44 A	(1·45)
C_1-C_2	1·54	(1·53)	C_6-C_7	1·51	(1·52)
C_2-C_3	1·31	(1·32)	C_7-C_8	1·31	(1·32)
C_3-C_4	1·53	(1·53)	C_8-C_9	1·53	(1·55)
C_3-C_5	1·51	(1·54)	C_8-C_{10}	1·54	(1·53)
N-C_1-C_2	109°	(109°)	C_3-C_5-C_6	115°	(112°)
C_1-C_2-C_3	126°	(126°)	C_5-C_6-C_7	112°	(112°)
C_2-C_3-C_4	124°	(124°)	C_6-C_7-C_8	129°	(129°)
C_2-C_3-C_5	121°	(119°)	C_7-C_8-C_9	124°	(123°)
C_4-C_3-C_5	115°	(117°)	C_7-C_8-C_{10}	121°	(121°)
			C_9-C_8-C_{10}	115°	(116°)

The general correctness of this solution is substantiated by the satisfactory agreement between the observed and calculated structure factors for all planes, and in place of the complete numerical list of 1060 values, which was submitted with the paper but not published, this agreement is shown diagrammatically in figure 3 for some 500 values.

In order to examine the accuracy of the results two further Fourier syntheses were evaluated. The full experimental data used previously contained a greater number of observed intensities for the *l* even series of planes than for the *l* odd by reason of the comparatively heavy chlorine atom contributing to the former but not the latter. It seemed advisable when locating the carbon atoms that the summations from the *l* even and *l* odd planes should include an equivalent number of terms. The synthesis was therefore repeated for the carbon skeleton omitting those

F values (c. 350) where the difference between F_{observed} and F_{chlorine} was less than 1·5 per molecule, this value being the minimum observable when F_{chlorine} was zero. The new atomic positions showed a mean shift of 0·02 A from those previously derived. The corresponding interatomic distances are given in parenthesis in table 1.

As an alternative method of balancing the l even and l odd terms, the F values were calculated for all planes of spacing greater than 0·9 A whose intensities were too small to be observed. These amounted to some 300 values which when included in the synthesis also resulted in shifts in the parameters of the order of 0·02 A and, as might be expected, considerably sharpened the electron peaks. The double bond lengths became 1·34 A, the C_5-C_6 bond 1·46 A, and the other bonds remained the same within 0·01 A. The progressive broadening of the contour intervals for each atom along the molecule was still clearly apparent, although the additional terms increased the maximum electron density of the peaks other than those at $y = \frac{1}{4}$ and $\frac{3}{4}$ by about $\frac{1}{2}$ electron/A^3.

Both these Fourier syntheses were an improvement on the original as regards the shape of the electron contours, particularly those of C_5 where the elongation of the peak in the c axis direction was completely removed. The changes in peak positions were of the same order as those between the penultimate and final successive Fourier syntheses, and the bond lengths and valency angles showed no variations beyond the limits of experimental errors which are estimated at ± 0·04 A and ± 4°.

Although the atom peaks were least sharp in the synthesis using the smaller number of F values, these atomic parameters (table 2) are considered to be the most reliable since they were derived solely from experimental data with the exclusion, for the organic part of the molecule, of such values as might impose a systematic error on the positions of the maxima.

TABLE 2

	x	y	z		x	y	z
Cl	0·053	0·250	0·080	C_5	0·253	0·625	0·182
N	0·049	0·750	0·216	C_6	0·304	0·700	0·287
C_1	0·103	0·750	0·333	C_7	0·357	0·746	0·207
C_2	0·158	0·751	0·255	C_8	0·394	0·916	0·212
C_3	0·198	0·593	0·264	C_9	0·447	0·918	0·118
C_4	0·198	0·379	0·358	C_{10}	0·392	1·120	0·315

THE STEREOCHEMISTRY OF THE MOLECULE

Within experimental error the two isoprene units are identical, the parameters of C_1-C_5 being more certain than C_6-C_{10} owing to the flattening of the Fourier peaks described above. They possess normal bond lengths ($C{=}C$ 1·34 A, $C{-}C$ 1·54 A) and the disposition of valencies about the double bonds is planar. Their angles appear to be distorted by about 5° from the theoretical 120° with the result that

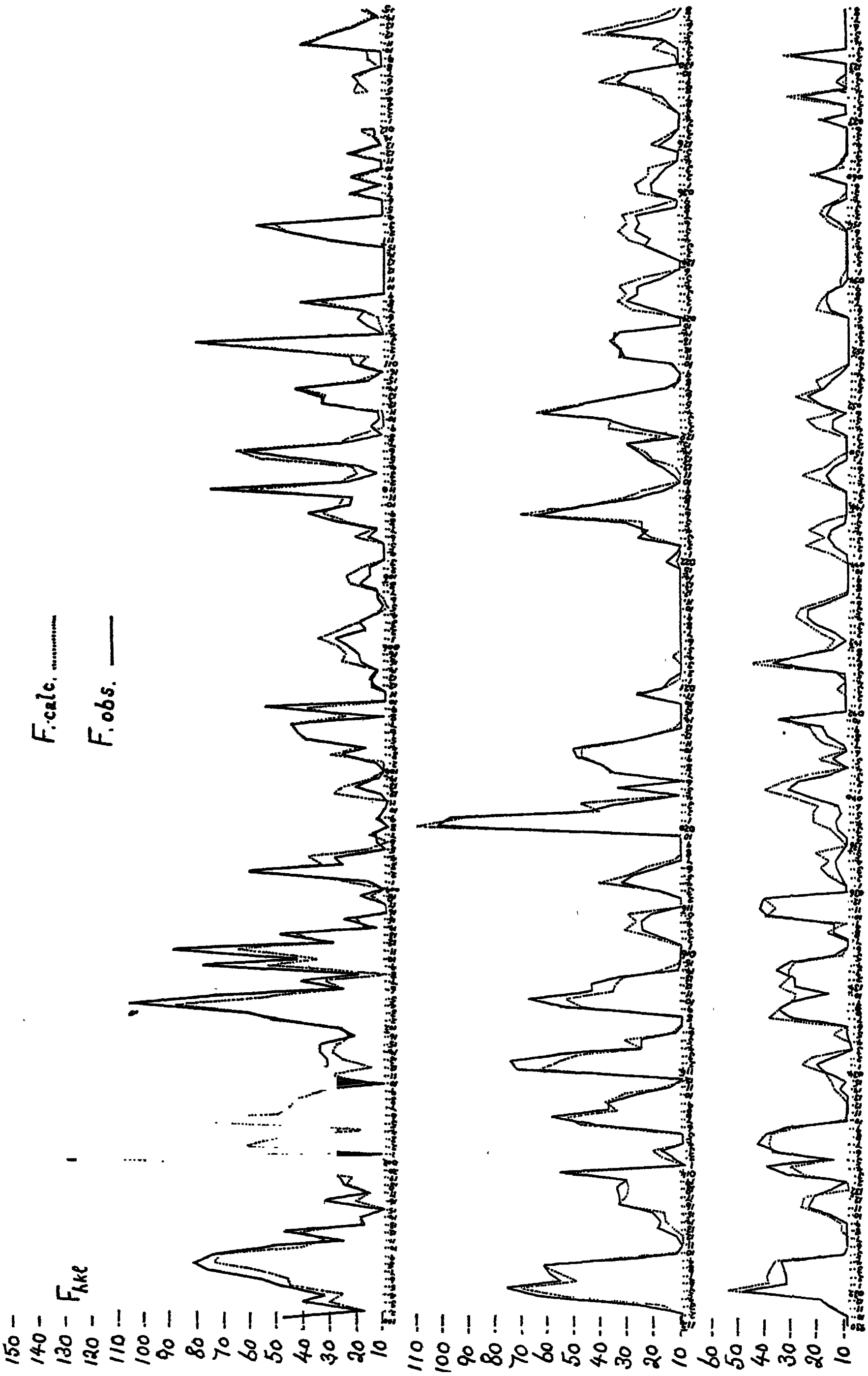
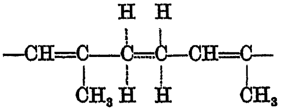

F. calc. ....
F. obs. ——
F_hkl
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
110
100
90
80
70
60
50
40
30
20
10
60
50
40
30
20
10

FIGURE 3

C_1-C_2 is not exactly parallel to C_3-C_5, nor C_6-C_7 to C_8-C_9. This variation, although small and very near the margin of error, may be real and characteristic of the unsymmetrical arrangement of the groups about the double bond.

The bond C_5-C_6 linking the isoprene units is shorter than that associated with a normal C-C bond by an amount two or three times the probable experimental error. Furthermore the distribution of the adjoining C_3-C_5 and C_6-C_7 is planar. These features, reminiscent of those found in conjugated molecules, seem inexplicable in terms of the usual theory of mesomerism. As reported by Bateman & Jeffrey (1943), the amine hydrochloride group does not appear to be directly responsible and the hybrid character may result from a hyperconjugation process in which the α-methylenic C-H electrons become partially localized in the central bond:



$$-\mathrm{CH}{=}\underset{\mathrm{CH_3}}{\mathrm{C}}-\overset{\mathrm{H}}{\underset{\mathrm{H}}{\mathrm{C}}}{=}\overset{\mathrm{H}}{\underset{\mathrm{H}}{\mathrm{C}}}-\mathrm{CH}{=}\underset{\mathrm{CH_3}}{\mathrm{C}}-$$

Evidence as to what extent this can be regarded as a general property of the polyisoprene system must await the examination of compounds with a similar structural modification.

Other stereochemical features of importance in the geranylamine structure concern the angular position of the C_5-C_6 bond relative to the two isoprene planes. Free rotation about the bond C_3-C_5 will be influenced by repulsion of the CH_2 group by the CH and CH_3. If the CH_2 group lies in the plane of the isoprene ($C_1 \ldots C_5$), then the approach between one or other pair of these atoms would be of the order of 2·8 A; a distance which in the absence of any special bonding between the hydrogen atoms provides an appreciable energy barrier. In the observed structure the C_5-C_6 bond is oriented at 80° to the isoprene plane and the steric repulsions balance at almost equal distances (C_6-C_2 3·28 A, C_6-C_4 3·24 A). In contrast, since there is no methyl addendum to C_7, the rotation about C_6-C_7 relative to the other isoprene plane is influenced solely by the approach of the CH_2 to the $CH_{3(10)}$. Over a large angular range, this repulsion will be ineffectual and the factors determining the observed angle of 45° between the C_5-C_6 bond and the $C_6 \ldots C_{10}$ plane must arise mainly from the packing of adjacent molecules in the crystal lattice, and probably from the relation between the terminal $C(CH_3)_2$ groups in particular. Whereas this latter steric effect is specific to the particular intermolecular arrangement, the former depends upon intramolecular forces and, since it will apply to any molecule with the same configuration, has an important bearing on the stereochemistry of the long-chain polymers.

THE INTERMOLECULAR ARRANGEMENT

The packing of the molecules in the crystal lattice is illustrated in figures 4 and 5. The hydrocarbon chains lie parallel with the molecules end to end in pairs. Each nitrogen is associated with four chlorine atoms at nearly equal distances.

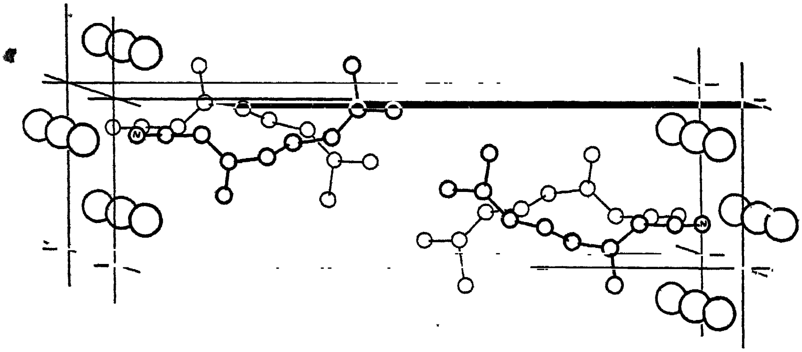


FIGURE 4. The structure of geranylamine hydrochloride.

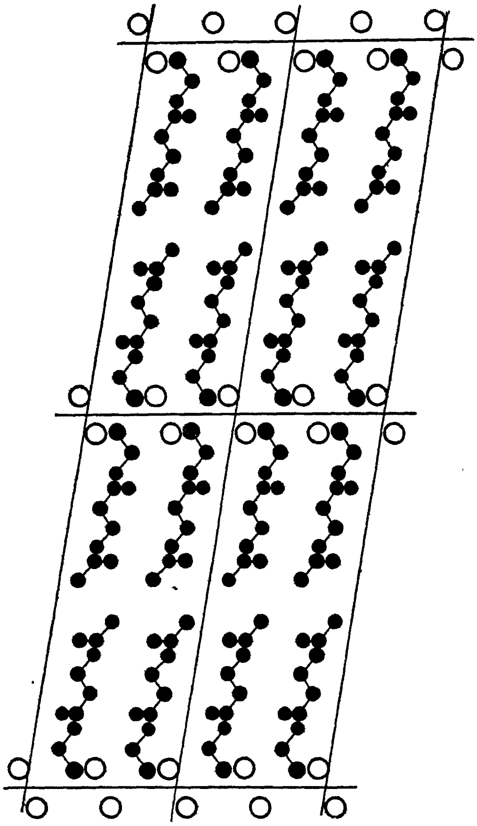


FIGURE 5. Projection of the structure on (010).

All other distances between atoms in adjacent molecules conform to Van der Waal's forces and are greater than 3·75 A with the exception of 3·60 A between centrosymmetrically related C_9 atoms.

TABLE 3

N-Cl_1	3·24 A	N-Cl_3	3·24 A
N-Cl_2	3·17	N-Cl_4	3·24

ACKNOWLEDGEMENTS

I am grateful to Professor Sir Lawrence Bragg and Dr E. G. Cox for advice and criticism and to Professor W. N. Haworth for facilities granted at Birmingham University for the earlier part of the work. Thanks are also given to Dr D. A. Sutton for preparation of the geranylamine hydrochloride.

The above work has been carried out as part of a programme of fundamental research undertaken by the Board of the British Rubber Producers' Research Association.

REFERENCES

Bateman & Jeffrey 1943 *Nature*, **152**, 446.
Beevers & Lipson 1936 *Proc. Phys. Soc.* **48**, 772.
Bunn 1942 *Proc. Roy. Soc.* A, **180**, 40.
Cox & Jeffrey 1942 *Trans. Faraday Soc.* **38**, 241.
Lotmar & Meyer 1936 *Mh. Chem.* **69**, 115.
Misch & Van der Wyk 1940 *J. Chem. Phys.* **8**, 127.
Morss 1938 *J. Amer. Chem. Soc.* **60**, 273.
Sauter 1937 *Z. Phys. Chem.* **36**B, 405.

For the tables of calculated and observed structure factors see pages 398–404.

Calculated and observed structure factors

hkl	$F_{\text{calc.}}$	$F_{\text{obs.}}$	*hkl*	$F_{\text{calc.}}$	$F_{\text{obs.}}$	*hkl*	$F_{\text{calc.}}$	$F_{\text{obs.}}$
100	48	48	$\bar{1}02$	26	27	$\bar{1}04$	5	<8
200	17	16	$\bar{2}02$	51	47	$\bar{2}04$	9	13
300	30	41	$\bar{3}02$	59	60	$\bar{3}04$	29	25
400	−25	33	$\bar{4}02$	87	106	$\bar{4}04$	18	16
500	45	50	$\bar{5}02$	54	76	$\bar{5}04$	39	38
600	−46	56	$\bar{6}02$	24	26	$\bar{6}04$	42	42
700	−55	76	$\bar{7}02$	32	41	$\bar{7}04$	43	45
800	−73	82	$\bar{8}02$	10	19	$\bar{8}04$	24	6
900	−71	74	$\bar{9}02$	53	78	$\bar{9}04$	−38	54
1000	−41	53	$\overline{10}02$	−35	42	$\overline{10}04$	6	6
1100	−32	25	$\overline{11}02$	−64	88	$\overline{11}04$	4	<8
1200	−42	47	$\overline{12}02$	−31	28	$\overline{12}04$	−12	13
1300	−17	16	$\overline{13}02$	−43	49	$\overline{13}04$	−14	13
1400	−18	18	$\overline{14}02$	−16	13	$\overline{14}04$	−13	12
1500	9	<8	$\overline{15}02$	−20	24	$\overline{15}04$	−25	18
1600	21	32	$\overline{16}02$	−8	7	$\overline{16}04$	−24	19
1700	14	15	$\overline{17}02$	−4	4	$\overline{17}04$	−14	12
1800	23	24	$\overline{18}02$	14	18	$\overline{18}04$	−17	14
1900	22	27	$\overline{19}02$	6	<8	1904	−1	<8
2000	41	47	$\overline{20}02$	−3	<8	2004	−1	5
2100	10	10	$\overline{21}02$	16	16	2104	6	7
2200	0	<8	$\overline{22}02$	12	13	2204	8	10
2300	3	<8	$\overline{23}02$	7	4	2304	10	10
2400	−8	7	$\overline{24}02$	13	10	2404	7	9
			$\overline{25}02$	4	<8	2504	6	8
002	−111	130						
102	19	10				006	−27	23
202	62	89	004	−1	<8	106	−34	27
302	−47	54	104	−26	16	206	−26	16
402	−18	27	204	−60	60	306	−18	18
502	−108	109	304	−33	25	406	−12	12
602	−49	66	404	−38	28	506	−4	10
702	−44	34	504	−16	<8	606	−8	13
802	−42	47	604	−4	12	706	12	12
902	−33	39	704	−15	13	806	22	18
1002	−15	<8	804	1	<8	906	17	14
1102	28	38	904	11	12	1006	23	18
1202	14	32	1004	9	<8	1106	18	16
1302	26	31	1104	−1	<8	1206	12	9
1402	29	34	1204	28	9	1306	5	<8
1502	27	34	1304	25	20	1406	5	<8
1602	23	21	1404	14	12	1506	−1	<8
1702	21	21	1504	13	8	1606	−1	<8
1802	13	9	1604	3	<8			
1902	1	<8	1704	5	<8	$\bar{1}06$	−23	14
2002	−13	12	1804	−2	<8	$\bar{2}06$	−18	14
2102	2	<8	1904	−2	<8	$\bar{3}06$	−5	<8
2202	3	<8	2004	−3	<8	$\bar{4}06$	0	<8
2302	−10	<8	2104	−9	<8	$\bar{5}06$	3	<8
2402	−5	10	2204	−10	8	$\bar{6}06$	20	15
2502	−14	11	2304	−5	6	$\bar{7}06$	12	12

CALCULATED AND OBSERVED STRUCTURE FACTORS

hkl	$F_{\text{calc.}}$	$F_{\text{obs.}}$	*hkl*	$F_{\text{calc.}}$	$F_{\text{obs.}}$	*hkl*	$F_{\text{calc.}}$	$F_{\text{obs.}}$
$\bar{8}06$	34	39	410	−21	18	$\overline{11}11$	4	<8
$\bar{9}06$	42	48	510	−56	39	$\overline{12}11$	−20	22
$\overline{10}06$	26	22	610	−65	62	$\overline{13}11$	2	<8
$\overline{11}06$	23	21	710	−26	25	$\overline{14}11$	11	12
$\overline{12}06$	15	10	810	−20	9	$\overline{15}11$	−1	<8
$\overline{13}06$	5	6	910	7	14	$\overline{16}11$	3	<8
$\overline{14}06$	−22	22	1010	6	11	$\overline{17}11$	15	14
$\overline{15}06$	−7	14	1110	10	11	$\overline{18}11$	7	8
$\overline{16}06$	−11	17	1210	33	33			
$\overline{17}06$	−14	16	1310	31	33	012	−18	13
$\overline{18}06$	−15	15	1410	40	43	112	−146	140
$\overline{19}06$	−17	16	1510	18	15	212	−79	84
$\overline{20}06$	−8	8	1610	6	<8	312	−59	60
$\overline{21}06$	−5	<8	1710	11	18	412	−47	43
$\overline{22}06$	−2	<8	1810	5	7	512	14	20
$\overline{23}06$	2	<8	1910	−15	16	612	−17	17
$\overline{24}06$	0	<8	2010	1	<8	712	−17	20
			2110	0	<8	812	11	<8
008	−12	6	2210	−12	12	912	14	9
108	3	5	2310	−8	8	1012	15	<8
208	−1	<8	2410	−8	8	1112	40	37
308	4	5				1212	29	29
408	10	8	011	18	21	1312	20	19
508	17	14	111	15	22	1412	11	11
608	24	16	211	49	50	1512	16	15
708	9	8	311	−80	80	1612	−7	9
808	4	12	411	8	<8	1712	−10	<8
908	13	12	511	10	15	1812	−15	<8
1008	2	<8	611	11	18	1912	−24	14
1108	6	12	711	16	16	2012	−7	8
			811	−33	41	2112	−8	9
$\bar{1}08$	−16	9	911	1	<8	2212	−9	<8
$\bar{2}08$	−17	12	1011	4	7	2312	−4	<8
$\bar{3}08$	−37	28	1111	4	<8	2412	−3	<8
$\bar{4}08$	−16	18	1211	4	<8	2512	4	<8
$\bar{5}08$	−10	<8	1311	−3	<8			
$\bar{6}08$	−7	<8	1411	−7	7	$\bar{1}12$	7	8
$\bar{7}08$	3	<8	1511	2	<8	$\bar{2}12$	5	<8
$\bar{8}08$	−4	<8	1611	0	<8	$\bar{3}12$	17	25
$\bar{9}08$	9	<8	1711	10	10	$\bar{4}12$	51	64
$\overline{10}08$	10	<8	1811	−4	<8	$\bar{5}12$	70	76
$\overline{11}08$	16	14	1911	−8	7	$\bar{6}12$	49	55
$\overline{12}08$	6	<8				$\bar{7}12$	55	59
$\overline{13}08$	10	10	$\bar{1}11$	−31	29	$\bar{8}12$	62	62
$\overline{14}08$	18	15	$\bar{2}11$	−44	46	$\bar{9}12$	37	32
$\overline{15}08$	8	8	$\bar{3}11$	51	58	$\overline{10}12$	11	11
$\overline{16}08$	2	<8	$\bar{4}11$	−3	<8	$\overline{11}12$	6	5
$\overline{17}08$	1	<8	$\bar{5}11$	5	<8	$\overline{12}12$	3	12
$\overline{18}08$	−1	<8	$\bar{6}11$	−4	5	$\overline{13}12$	−13	18
			$\bar{7}11$	−22	20	$\overline{14}12$	−15	17
110	60	75	$\bar{8}11$	0	<8	$\overline{15}12$	−16	20
210	−26	23	$\bar{9}11$	17	21	$\overline{16}12$	−33	33
310	−20	12	$\overline{10}11$	−3	<8	$\overline{17}12$	−30	33

CALCULATED AND OBSERVED STRUCTURE FACTORS

hkl	$F_{\text{calc.}}$	$F_{\text{obs.}}$
$\overline{18}12$	−29	34
$\overline{19}12$	−30	30
$\overline{20}12$	−1	<8
$\overline{21}12$	3	4
$\overline{22}12$	2	6
$\overline{23}12$	7	9
$\overline{24}12$	7	12
$\overline{25}12$	11	13
013	−14	11
113	−6	<8
213	−9	8
313	18	17
413	7	8
513	−17	18
613	4	9
713	−10	8
813	22	20
913	−4	<8
1013	−8	12
$\overline{1}13$	10	10
$\overline{2}13$	5	5
$\overline{3}13$	4	8
$\overline{4}13$	−3	<8
$\overline{5}13$	0	<8
$\overline{6}13$	−14	15
$\overline{7}13$	12	17
$\overline{8}13$	9	13
$\overline{9}13$	−8	11
$\overline{10}13$	−8	8
$\overline{11}13$	−5	6
$\overline{12}13$	18	17
$\overline{13}13$	4	<8
$\overline{14}13$	−1	<8
$\overline{15}13$	−5	6
$\overline{16}13$	−1	<8
$\overline{17}13$	7	9
014	−53	55
114	−2	4
214	−16	12
314	−20	17
414	7	<8
514	−7	<8
614	47	43
714	52	59
814	33	36
914	38	36
1014	31	26
1114	16	13
1214	1	<8
1314	0	<8
1414	−10	10
1514	−18	14
1614	−16	14
1714	−15	14
1814	−15	12
1914	−12	9
$\overline{1}14$	−63	72
$\overline{2}14$	−56	74
$\overline{3}14$	−48	56
$\overline{4}14$	−24	24
$\overline{5}14$	−31	24
$\overline{6}14$	−13	12
$\overline{7}14$	8	<8
$\overline{8}14$	2	<8
$\overline{9}14$	46	53
$\overline{10}14$	53	67
$\overline{11}14$	35	43
$\overline{12}14$	34	43
$\overline{13}14$	31	33
$\overline{14}14$	18	11
$\overline{15}14$	−7	9
$\overline{16}14$	−5	8
$\overline{17}14$	−13	14
$\overline{18}14$	−13	13
$\overline{19}14$	−9	11
$\overline{20}14$	−21	21
$\overline{21}14$	−16	16
$\overline{22}14$	−10	10
$\overline{23}14$	−6	<8
015	−5	<8
115	4	<8
215	−3	<8
315	4	<8
415	−7	<8
515	1	<8
615	0	<8
715	−1	<8
815	−3	<8
$\overline{1}15$	2	<8
$\overline{2}15$	1	<8
$\overline{3}15$	−3	<8
$\overline{4}15$	−3	<8
$\overline{5}15$	−2	<8
$\overline{6}15$	12	8
$\overline{7}15$	−9	8
$\overline{8}15$	0	<8
$\overline{9}15$	−2	<8
$\overline{10}15$	5	8
016	2	<8
116	−1	<8
216	18	15
316	30	24
416	27	25
516	29	21
616	8	<8
716	3	<8
816	14	<8
916	3	<8
1016	−6	6
1116	−7	6
1216	−5	6
1316	−13	9
1416	−13	10
1516	−11	11
1616	−12	8
$\overline{1}16$	−6	<8
$\overline{2}16$	−14	<8
$\overline{3}16$	−17	17
$\overline{4}16$	−40	31
$\overline{5}16$	−28	26
$\overline{6}16$	−22	18
$\overline{7}16$	−8	<8
$\overline{8}16$	−10	<8
$\overline{9}16$	−8	<8
$\overline{10}16$	−5	<8
$\overline{11}16$	10	<8
$\overline{12}16$	13	8
$\overline{13}16$	10	8
$\overline{14}16$	16	16
$\overline{15}16$	20	17
$\overline{16}16$	14	10
$\overline{17}16$	9	8
$\overline{18}16$	4	<8
$\overline{19}16$	−2	<8
$\overline{20}16$	−4	<8
$\overline{21}16$	−3	<8
$\overline{22}16$	−5	<8
$\overline{23}16$	−5	7
$\overline{24}16$	−11	8
018	8	8
118	11	9
218	10	8
318	13	9
418	9	7
518	6	7
618	8	7
718	6	6

CALCULATED AND OBSERVED STRUCTURE FACTORS

hkl	$F_{\text{calc.}}$	$F_{\text{obs.}}$
$\bar{1}18$	8	<8
$\bar{2}18$	5	<8
$\bar{3}18$	−9	<8
$\bar{4}18$	2	<8
$\bar{5}18$	−9	<8
$\bar{6}18$	−10	8
$\bar{7}18$	−16	14
$\bar{8}18$	−24	17
$\bar{9}18$	−12	9
$\overline{10}18$	−3	<8
$\overline{11}18$	−5	<8
$\overline{12}18$	2	<8
$\overline{13}18$	8	9
$\overline{14}18$	9	9
$\overline{15}18$	9	10
$\overline{16}18$	7	<8
$\overline{17}18$	10	9
$\overline{18}18$	2	<8
020	−110	103
120	−92	97
220	−40	43
320	−47	34
420	8	8
520	33	32
620	4	10
720	36	34
820	39	39
920	47	49
1020	47	50
1120	32	32
1220	20	20
1320	12	8
1420	0	<8
1520	8	8
1620	−9	9
1720	−25	26
1820	−19	24
1920	−28	30
021	−15	17
121	−10	14
221	5	<8
321	8	<8
421	−13	8
521	4	<8
621	−6	9
721	−4	8
821	−3	<8
921	3	<8
1021	−4	9
1121	3	<8
1221	6	<8
1321	−2	<8
1421	8	8
1521	−2	<8
1621	−2	<8
1721	2	<8
1821	7	<8
1921	−1	<8
2021	−4	<8
2121	2	<8
$\bar{1}21$	−4	<8
$\bar{2}21$	33	35
$\bar{3}21$	−34	37
$\bar{4}21$	−8	<8
$\bar{5}21$	4	8
$\bar{6}21$	12	13
$\bar{7}21$	4	8
$\bar{8}21$	−13	13
$\bar{9}21$	5	<8
$\overline{10}21$	−2	<8
$\overline{11}21$	0	<8
$\overline{12}21$	0	<8
$\overline{13}21$	1	<8
$\overline{14}21$	−2	<8
$\overline{15}21$	1	<8
$\overline{16}21$	−8	11
022	11	15
122	−7	11
222	0	<8
322	21	25
422	24	21
522	24	30
622	70	64
722	44	46
822	29	27
922	24	16
1022	−5	9
1122	−6	12
1222	−12	16
1322	−20	21
1422	−23	25
1522	−29	29
1622	−22	20
1722	−10	7
1822	−8	<8
1922	0	<8
2022	5	<8
$\bar{1}22$	−16	7
$\bar{2}22$	−36	27
$\bar{3}22$	−35	37
$\bar{4}22$	−63	61
$\bar{5}22$	−45	45
$\bar{6}22$	−30	32
$\bar{7}22$	−14	15
$\bar{8}22$	−8	<8
$\bar{9}22$	8	<8
$\overline{10}22$	13	12
$\overline{11}22$	33	30
$\overline{12}22$	32	32
$\overline{13}22$	34	35
$\overline{14}22$	34	32
$\overline{15}22$	0	<8
$\overline{16}22$	7	<8
$\overline{17}22$	1	<8
$\overline{18}22$	−6	<8
$\overline{19}22$	−2	<8
$\overline{20}22$	−6	6
$\overline{21}22$	−10	9
$\overline{22}22$	−10	8
$\overline{23}22$	−7	9
023	19	17
123	3	<8
223	8	<8
323	−31	25
423	10	8
523	9	13
623	4	10
723	2	<8
823	−16	11
923	11	11
$\bar{1}23$	−2	<8
$\bar{2}23$	−32	29
$\bar{3}23$	10	<8
$\bar{4}23$	8	8
$\bar{5}23$	0	<8
$\bar{6}23$	0	<8
$\bar{7}23$	−5	<8
$\bar{8}23$	−3	<8
$\bar{9}23$	1	<8
$\overline{10}23$	0	<8
$\overline{11}23$	14	15
$\overline{12}23$	−15	13
$\overline{13}23$	−4	<8
$\overline{14}23$	5	<8
$\overline{15}23$	1	<8
$\overline{16}23$	8	8
$\overline{17}23$	−13	12
024	2	<8
124	27	18
224	33	29
324	29	25

CALCULATED AND OBSERVED STRUCTURE FACTORS

hkl	$F_{\text{calc.}}$	$F_{\text{obs.}}$
424	32	25
524	24	16
624	1	<8
724	1	<8
824	7	<8
924	−10	<8
1024	1	<8
1124	−9	8
1224	−23	13
1324	−17	11
1424	−15	9
1524	−14	8
1624	−4	<8
1724	−4	<8
1824	0	<8
1924	2	<8
$\bar{1}24$	1	<8
$\bar{2}24$	−15	16
$\bar{3}24$	−21	22
$\bar{4}24$	−28	20
$\bar{5}24$	−32	26
$\bar{6}24$	−29	24
$\bar{7}24$	−31	26
$\bar{8}24$	−19	10
$\bar{9}24$	0	<8
$\overline{10}24$	7	9
$\overline{11}24$	−5	<8
$\overline{12}24$	15	15
$\overline{13}24$	4	<8
$\overline{14}24$	9	<8
$\overline{15}24$	27	21
$\overline{16}24$	18	15
$\overline{17}24$	15	12
$\overline{18}24$	10	8
$\overline{19}24$	3	<8
$\overline{20}24$	3	<8
$\overline{21}24$	−3	<8
$\overline{22}24$	−7	9
$\overline{23}24$	−7	8
$\overline{24}24$	−9	9
$\overline{25}24$	−8	8
025	−7	7
125	0	<8
225	−4	<8
325	12	8
425	−5	<8
525	−12	8
625	2	<8
725	5	6
825	12	8
$\bar{1}25$	−8	8
$\bar{2}25$	4	<8
$\bar{3}25$	7	<8
$\bar{4}25$	−4	<8
$\bar{5}25$	0	<8
$\bar{6}25$	−16	9
$\bar{7}25$	5	<8
$\bar{8}25$	7	<8
$\bar{9}25$	3	<8
$\overline{10}25$	−2	<8
$\overline{11}25$	−18	9
$\overline{12}25$	−14	11
026	24	19
126	26	20
226	21	17
326	21	14
426	10	10
526	7	7
626	4	<8
726	−7	<8
826	−20	14
926	−22	17
1026	−23	21
1126	−21	17
1226	2	<8
1326	−7	<8
$\bar{1}26$	19	17
$\bar{2}26$	14	12
$\bar{3}26$	4	<8
$\bar{4}26$	2	<8
$\bar{5}26$	−6	6
$\bar{6}26$	−14	13
$\bar{7}26$	−19	16
$\bar{8}26$	−20	24
$\bar{9}26$	−38	32
$\overline{10}26$	−28	28
$\overline{11}26$	−16	17
$\overline{12}26$	−16	8
$\overline{13}26$	5	8
$\overline{14}26$	8	11
$\overline{15}26$	8	9
$\overline{16}26$	10	9
$\overline{17}26$	13	12
$\overline{18}26$	12	13
$\overline{19}26$	12	12
$\overline{20}26$	6	7
028	10	8
128	2	<8
228	−2	<8
328	−6	6
428	−9	7
528	−12	9
628	−12	12
728	−11	9
$\bar{1}28$	12	8
$\bar{2}28$	18	13
$\bar{3}28$	21	15
$\bar{4}28$	18	13
$\bar{5}28$	9	8
$\bar{6}28$	8	7
$\bar{7}28$	−1	<8
$\bar{8}28$	−10	17
$\bar{9}28$	6	8
$\overline{10}28$	−6	<8
$\overline{11}28$	−8	8
$\overline{12}28$	−8	8
$\overline{13}28$	−13	11
$\overline{14}28$	−10	9
$\overline{15}28$	−6	<8
$\overline{16}28$	−4	<8
$\overline{17}28$	−2	<8
$\overline{18}28$	−4	<8
130	1	<8
230	11	8
330	18	10
430	15	15
530	45	37
630	30	21
730	17	11
830	10	8
930	4	<8
1030	−4	<8
1130	−16	15
1230	−18	18
1330	−21	18
1430	−25	23
1530	−25	20
1630	−2	<8
031	−3	<8
131	−14	13
231	−3	<8
331	13	9
531	−4	<8
431	−3	<8
631	−2	<8
731	−4	<8
831	9	11
931	2	<8
1031	−4	<8
1131	0	<8
1231	−1	<8
$\bar{1}31$	8	<8
$\bar{2}31$	5	<8

CALCULATED AND OBSERVED STRUCTURE FACTORS

hkl	$F_{\text{calc.}}$	$F_{\text{obs.}}$	*hkl*	$F_{\text{calc.}}$	$F_{\text{obs.}}$	*hkl*	$F_{\text{calc.}}$	$F_{\text{obs.}}$
$\bar{3}31$	14	12	$\overline{20}32$	10	9	$\bar{1}36$	8	8
$\bar{4}31$	−3	<6	$\overline{21}32$	0	<8	$\bar{2}36$	11	9
$\bar{5}31$	−4	<6				$\bar{3}36$	15	13
$\bar{6}31$	−2	<6	034	17	17	$\bar{4}36$	21	13
$\bar{7}31$	5	7	134	25	20	$\bar{5}36$	22	11
$\bar{8}31$	6	5	234	21	13	$\bar{6}36$	19	12
$\bar{9}31$	−4	<8	334	0	<8	$\bar{7}36$	5	8
$\overline{10}31$	−3	<8	434	−7	7	$\bar{8}36$	1	<8
$\overline{11}31$	0	<8	534	−14	7	$\bar{9}36$	1	<8
$\overline{12}31$	7	<8	634	−4	<8	$\overline{10}36$	−5	<8
$\overline{13}31$	2	<8	734	−38	29	$\overline{11}36$	4	<8
$\overline{14}31$	−2	<8	834	−34	34	$\overline{12}36$	−5	<8
$\overline{15}31$	−2	<8	934	−29	29	$\overline{13}36$	−10	8
			1034	−28	16	$\overline{14}36$	−16	10
032	54	48	1134	2	<8	$\overline{15}36$	−15	10
132	38	31	1234	−6	<8	$\overline{16}36$	−5	<8
232	39	31	1334	−3	<8			
332	38	32	1434	4	<8	038	−3	<8
432	36	34	1534	6	<8	138	−4	<8
532	10	10	1634	13	11	$\bar{1}38$	−2	<8
632	−9	8	1734	15	1[illegible]	$\bar{2}38$	2	<8
732	9	8	1834	8	6	$\bar{3}38$	−2	<8
832	−11	8				$\bar{4}38$	4	<8
932	−11	8	$\bar{1}34$	33	28	$\bar{5}38$	8	8
1032	−17	16	$\bar{2}34$	30	28	$\bar{6}38$	10	8
1132	−26	23	$\bar{3}34$	30	35	$\bar{7}38$	3	<8
1232	−25	21	$\bar{4}34$	27	30	$\bar{8}38$	12	8
1332	−18	17	$\bar{5}34$	11	13	$\bar{9}38$	13	10
1432	−12	8	$\bar{6}34$	1	<8	$\overline{10}38$	5	10
1532	−2	<8	$\bar{7}34$	−6	7	$\overline{11}38$	3	8
1632	−7	<8	$\bar{8}34$	−16	14			
1732	−4	<8	$\bar{9}34$	−8	14	040	29	23
1832	7	<8	$\overline{10}34$	−39	39	140	39	30
1932	6	<8	$\overline{11}34$	−36	42	240	31	27
			$\overline{12}34$	−29	30	340	22	15
$\bar{1}32$	9	15	$\overline{13}34$	−26	20	440	−4	<8
$\bar{2}32$	7	<8	$\overline{14}34$	9	<8	540	19	14
$\bar{3}32$	−24	31	$\overline{15}34$	4	<8	640	−1	<8
$\bar{4}32$	−31	39	$\overline{16}34$	−2	<8	740	−17	9
$\bar{5}32$	−13	15	$\overline{17}34$	7	9	840	−20	16
$\bar{6}32$	−37	34	$\overline{18}34$	8	9	940	−22	17
$\bar{7}32$	−38	42	$\overline{19}34$	13	13	1040	−33	34
$\bar{8}32$	−37	39	$\overline{20}34$	9	9	1140	−21	22
$\bar{9}32$	−26	31	$\overline{21}34$	7	<8	1240	−13	18
$\overline{10}32$	−10	8				1340	−6	8
$\overline{11}32$	−4	<8	036	1	<8			
$\overline{12}32$	−4	<8	136	−6	6	042	9	8
$\overline{13}32$	11	8	236	−16	11	142	−2	<8
$\overline{14}32$	−13	12	336	−18	13	242	−6	10
$\overline{15}32$	16	8	436	−16	14	342	−13	9
$\overline{16}32$	28	26	536	−20	12	442	−9	8
$\overline{17}32$	16	14	636	−15	<8	542	−10	13
$\overline{18}32$	21	23	736	−5	<8	642	−44	36
$\overline{19}32$	10	19	836	−3	<8	742	−28	28

CALCULATED AND OBSERVED STRUCTURE FACTORS

hkl	$F_{\text{calc.}}$	$F_{\text{obs.}}$
842	−16	14
942	−10	8
1042	3	<8
1142	−3	<8
1242	6	<8
$\bar{1}42$	14	9
$\bar{2}42$	18	14
$\bar{3}42$	23	20
$\bar{4}42$	27	24
$\bar{5}42$	26	24
$\bar{6}42$	18	16
$\bar{7}42$	10	8
$\bar{8}42$	3	<8
$\bar{9}42$	−10	8
$\overline{10}42$	−8	8
$\overline{11}42$	−13	10
$\overline{12}42$	−19	13
044	−2	<8
144	−14	9
244	−23	14
344	−17	16
444	−22	14
544	−17	8
644	2	<8
744	−1	<8
844	−4	<8
944	2	<8
1044	0	<8
$\bar{1}44$	−2	<8
$\bar{2}44$	8	7
$\bar{3}44$	7	9
$\bar{4}44$	18	15
$\bar{5}44$	25	16
$\bar{6}44$	16	14
$\bar{7}44$	13	8
$\bar{8}44$	6	8
$\bar{9}44$	6	8
$\overline{10}44$	−2	<8
$\overline{11}44$	−4	<8
$\overline{12}44$	−7	8
046	−8	12
146	−13	11
246	−10	8
$\bar{1}46$	−13	11
$\bar{2}46$	−9	9
$\bar{3}46$	−9	7
$\bar{4}46$	0	<8
$\bar{5}46$	7	<8
$\bar{6}46$	5	<8
$\bar{7}46$	8	14
$\bar{8}46$	10	16

hkl	$F_{\text{calc.}}$	$F_{\text{obs.}}$
041	5	<8
141	−5	<8
241	0	<8
341	9	9
441	−6	<8
541	−1	<8
$\bar{1}41$	−8	<8
$\bar{2}41$	4	<8
$\bar{3}41$	0	<8
$\bar{4}41$	2	<8
150	4	<8
250	−10	8
350	−7	8
450	−20	14
550	−23	14
650	−13	9
750	−9	8
850	−7	8
950	−4	<8
1050	1	<8
1150	8	<8
052	−18	18
152	−28	23
252	−18	19
352	−20	15
452	−14	11
552	−3	<8
652	−4	<8
752	1	<8
852	3	<8
$\bar{1}52$	−8	<8
$\bar{2}52$	−2	<8
$\bar{3}52$	8	8
$\bar{4}52$	7	8
$\bar{5}52$	8	9
$\bar{6}52$	20	15
$\bar{7}52$	18	17
$\bar{8}52$	16	16
$\bar{9}52$	11	11
$\overline{10}52$	2	<8
$\overline{11}52$	6	<8
054	−11	8
154	−8	8
254	−9	8
354	−5	<8
454	4	<8
554	1	<8
654	8	<8
754	15	10

hkl	$F_{\text{calc.}}$	$F_{\text{obs.}}$
854	16	14
954	14	12
1054	10	20
$\bar{1}54$	−15	12
$\bar{2}54$	−18	16
$\bar{3}54$	−16	15
$\bar{4}54$	−11	9
$\bar{5}54$	−5	<8
$\bar{6}54$	−6	<8
$\bar{7}54$	4	<8
$\bar{8}54$	4	<8
$\bar{9}54$	8	<8
060	−22	20
160	−11	14
260	−11	10
360	−5	6
460	3	<8
560	−3	<8
660	4	<8
760	8	9
860	11	11
960	10	13
1060	9	11
062	4	<8
162	−1	<8
262	−4	<8
362	6	<8
462	0	<8
562	13	9
662	13	9
762	8	8
$\bar{1}62$	−7	<8
$\bar{2}62$	−7	<8
$\bar{3}62$	−13	10
$\bar{4}62$	−14	11
$\bar{5}62$	−7	7
064	−1	<8
164	2	<8
264	12	8
364	4	<8
464	5	<8
$\bar{1}64$	−2	<8
$\bar{2}64$	−3	<8
$\bar{3}64$	−3	<8
$\bar{4}64$	−1	<8
$\bar{5}64$	−11	8
$\bar{6}64$	−7	6
$\bar{7}64$	−4	<8

The statistical theory of the strength of bundles of threads. I

By H. E. Daniels, Ph.D., *Wool Industries Research Association*

(*Communicated by H. Jeffreys, F.R.S.—Received* 31 *May* 1944)

A group of parallel threads of equal length, clamped at each end so that all threads extend equally under tension, is called a bundle, and the maximum load which the bundle can support is called its strength. The object of the work is to study the probability distribution of the strength of bundles whose constituent threads are sampled randomly from an infinite population of threads in which the probability distribution of strength is known.

The relation between the strength of a bundle and the strengths of its constituent threads is first discussed, and results are stated for bundles so large that the proportions of threads of different strengths approach their expectations. The properties of the probability distribution of bundle strength are next developed in detail, attention being confined in the present paper to the case where all threads have the same load-extension curve up to breaking point. Finally, the asymptotic behaviour of the distribution for large numbers of threads is studied, and it is shown that in the commonest cases the distribution tends to assume the normal form.

Introduction

1. Consider an infinite population of threads of equal length whose probability distribution of strength is known. A number of threads are sampled randomly from this population, laid side by side and clamped at each end to form what we term a *bundle*. When a free load is applied to the bundle all threads extend by an equal amount, and the minimum load beyond which all the threads of the bundle break is called its *strength*. The present work is an investigation into the probability distribution of the strength of such bundles.

The problem presents itself naturally in the theory of strength testing of textile materials. For example, a test widely used in practice for estimating the strength of yarns is known as the 'hank' test for wool yarns, and the 'lea' test for cotton yarns. The method employed is to reel a hank of yarn containing a stated number of turns of specified circumference and to apply the breaking load to the hank by stretching it between two hooks. The word 'bundle' has been used here in preference to 'hank' or 'lea', since there is usually a small amount of slip at the hooks which makes the conditions of this test indeterminate. The testing of cloth samples for warp and weft strength is another instance of a practical test in which the specimens approximate to bundles, though in closely woven fabrics the cross-threads afford a measure of support which introduces complications. Other examples outside the field of textile testing could no doubt also be quoted.

2. The subject was first considered, as far as the author is aware, by F. T. Peirce (1926), relevant experimental work being published in previous papers in the same series. He deals exhaustively with the underlying physical considerations and derives useful formulae for the strength of large bundles.

The wider significance of the problem was also recognized by Peirce. He points out that a study of the strength properties of certain materials must involve considerations fundamentally similar to those arising in the theory of bundles (called by Peirce 'composite specimens'), since each element of the material may be thought of as made up of subelements arranged both in series and parallel along a particular direction of stress. Recently a notable attempt has been made by W. Weibull (1939) to evolve a statistical theory of strength of materials, in the course of which consideration is given to the behaviour under a free load of materials composed of independent parallel elements (the case of so-called 'incoherent irregularity')*.

The present paper opens with a preliminary section on the relation between the strength of a bundle and the strengths of its constituent threads, in which the work overlaps that of Peirce to some extent. The distribution which forms the main topic of the paper was not, however, discussed by Peirce. Apart from its practical application, it is of considerable inherent interest and appears hitherto to have received little attention from mathematicians.

THE RELATION BETWEEN A BUNDLE AND ITS CONSTITUENT THREADS

3. It is assumed that for each thread there is a definite extension and load at which it breaks, so that threads which yield rather than snap are excluded from the discussion. Whatever assumptions are made about the elastic properties of the threads, the following sequence of events occurs when a given load is applied. Suppose the load is increased gradually from zero to its final value. At first it is distributed in some way between the n threads, and equilibrium may be reached without any thread giving way. If the load is great enough, however, one of the threads breaks at some stage and the load is redistributed among the remaining $n-1$ threads, each bearing a somewhat larger share of the load than before and so being more likely to give way. Similarly, a number of threads may break successively until either a point is reached where the remaining threads have sufficient strength to maintain between them the final load, or no such point is reached and all the threads ultimately give way, in which case, of course, the bundle is broken.

To formulate the problem more precisely, it is necessary to know how the load distributes itself between the individual threads of the bundle, and this depends on the elastic properties of the threads. It should be observed that if they are assumed to be inextensible the problem has no definite solution. A familiar example of the same state of affairs is that of a rigid beam supported horizontally at more than two points, in which case the pressures on the supports cannot be deduced.

---

* Unfortunately there appears to be a flaw in his discussion of probabilities, and formulae are obtained which are open to question. I suggest that his equations (136) and (137) should read

$$dS'_{2/2} = 2[S(2\sigma) - S(\sigma)]\,dS(\sigma)$$

and

$$dS''_{2/2} = 2S(\sigma)\,dS(2\sigma)$$

respectively. The final expression for $S_{2/2} = S'_{2/2} + S''_{2/2}$ then agrees with the present formula (9·1) for $n = 2$.

Denoting load by S and extension by e, the load-extension relation for a given thread can be written

$$S = f(e, \alpha),$$

where α represents, for simplicity, one parameter defining the curve specific to a particular thread. The value of α is in the general case correlated with the breaking extension ϵ, or the breaking load $\sigma = f(\epsilon, \alpha)$.

Large bundles

4. It is convenient first to consider very large bundles within which the actual numbers of threads with given properties can be treated as equal to their expectations. (See, however, the remark at the end of § 7.)

The probability of ϵ and α lying in specified small intervals is denoted by

$$\phi(\epsilon, \alpha)\, d\alpha\, d\epsilon \tag{4·1}$$

and the probability density for ϵ for each thread is

$$\psi(\epsilon) = \int_0^\infty \phi(\epsilon, \alpha)\, d\alpha. \tag{4·2}$$

Let a load S be applied to the bundle, and let the corresponding extension be e. If the bundle does not break, a state of equilibrium is set up in which there are r surviving threads out of the original n, where r satisfies the relation

$$\frac{r}{n} = \int_e^\infty \psi(\epsilon)\, d\epsilon = \int_e^\infty d\epsilon \int_0^\infty \phi(\epsilon, \alpha)\, d\alpha. \tag{4·3}$$

At extension e the load on a particular thread is $f(e, \alpha)$, and when equilibrium is established, the total load is distributed over the surviving threads so as to make

$$S = r\frac{\int_e^\infty d\epsilon \int_0^\infty f(e, \alpha)\, \phi(\epsilon, \alpha)\, d\alpha}{\int_e^\infty d\epsilon \int_0^\infty \phi(\epsilon, \alpha)\, d\alpha}. \tag{4·4}$$

Hence from (4·3) and (4·4) the load-extension relation for the bundle is

$$S = n\int_e^\infty d\epsilon \int_0^\infty f(e, \alpha)\, \phi(\epsilon, \alpha)\, d\alpha. \tag{4·5}$$

Breaking extension e_T is attained when S has its greatest possible value S_T, and e_T is therefore the root of the equation

$$\frac{d}{de_T}\int_{e_T}^\infty d\epsilon \int_0^\infty f(e_T, \alpha)\, \phi(\epsilon, \alpha)\, d\alpha = 0, \tag{4·6}$$

which gives S its greatest value.

It is worth noting from (4·5) that the load-extension curve is similar for all large bundles, the load for a given extension being directly proportional to the original number of constituent threads.

5. We now introduce certain restrictions on the behaviour of the threads which lead to simplifications in the theory.

(i) In many practical cases the load-extension curves for all threads are similar in form, differing only in a scale factor. This might be expected, for example, when the parameter α is the thickness of the thread. Then

$$f(e,\alpha) = \alpha f(e) \tag{5·1}$$

and (4·5) becomes
$$S = nf(e)\int_e^\infty \alpha(\epsilon)\,\phi(\epsilon,\alpha)\,d\epsilon, \tag{5·2}$$

where $\alpha(\epsilon)$ is the expectation of α given that the thread has breaking extension ϵ.

The function $f(e)$ may reasonably be taken to increase steadily with e, and no generality is then sacrificed if Hooke's law is assumed, since otherwise we may always transform to a new 'extension' $f(e)$. The load-extension curve is therefore taken to be

$$f(e,\alpha) = \alpha e \tag{5·3}$$

and (5·2) becomes
$$S = ne\int_e^\infty \alpha(\epsilon)\,\phi(\epsilon,\alpha)\,d\epsilon. \tag{5·4}$$

(ii) It is found experimentally (cf. Peirce 1926) that the probabilities of α and ϵ are often distributed independently, in which case $\phi(\epsilon,\alpha) = \psi(\epsilon)\phi(\alpha)$ and $\alpha(\epsilon) = \bar{\alpha}$, independent of ϵ. Then (5·4) becomes

$$S = n\bar{\alpha}e\int_e^\infty \psi(\epsilon)\,d\epsilon \tag{5·5}$$

and the breaking extension of the bundle satisfies

$$\frac{d}{de_r}\left\{e_r\int_{e_r}^\infty \psi(\epsilon)\,d\epsilon\right\} = 0. \tag{5·6}$$

(iii) Finally, the dispersion of α may be negligibly small compared with that of breaking extension ϵ. That this is a reasonable assumption in certain circumstances is evident on considering a thread composed of a large number N of independent consecutive elements, such as, for example, a long chain, the rth element obeying Hooke's law with $\alpha = \alpha_r$. Under a load s each element extends by $e_r = s/\alpha_r$ and the total extension is

$$e = \frac{1}{N}\sum_{r=1}^{N} e_r = \frac{s}{N}\sum_{r=1}^{N}\frac{1}{\alpha_r}.$$

The load-extension ratio for the whole thread is thus α, where

$$\frac{1}{\alpha} = \frac{1}{N}\sum_{r=1}^{N}\frac{1}{\alpha_r},$$

and it follows that as the number N of elements is increased the standard deviation of α decreases as $N^{-\frac{1}{2}}$.

But if $\beta(s)$ is the chance of an element succumbing under load s, the chance of the whole thread giving way is

$$b(s) = 1-[1-\beta(s)]^N.$$

When N is large, and $\beta(s) \sim \kappa s^\gamma$ $(\gamma > 0)$ near $s = 0$, the limiting form of $b(s)$ is

$$b(s) \sim 1-e^{-N\kappa s^\gamma}. \qquad (5{\cdot}7)$$

Hence the expectation of s and the standard deviation of s are both proportional to $(N\kappa)^{1/\gamma}$ and so have a ratio independent of N.

We have reduced the problem to the simplest case when all threads have the same load-extension ratio and each thread bears an equal share of the load. Writing $\theta(\sigma)$ for the probability density of breaking load σ, where

$$\theta(\sigma) = \frac{1}{\alpha}\psi\left(\frac{\sigma}{\alpha}\right),$$

the relation between the total load S and the load s on each surviving thread at equilibrium is

$$S = ns\int_s^\infty \theta(\sigma)\,d\sigma, \qquad (5{\cdot}8)$$

and the breaking load S_τ of the bundle occurs when $s = s_\tau$ satisfies

$$\frac{d}{ds_\tau}\left\{s_\tau\int_{s_\tau}^\infty \theta(\sigma)\,d\sigma\right\} = 0, \qquad (5{\cdot}9)$$

and gives S its greatest value S_τ.

6. A very useful way of representing these results graphically is now introduced. The notation

$$\int_0^s \theta(\sigma)\,d\sigma = b(s) = a\left(\frac{1}{s}\right) = a(w) \qquad (6{\cdot}1)$$

is used for the chance of a thread not exceeding s in strength, where w may be called the 'weakness' of the thread. Then (5·8) and (5·9) become

$$S = \frac{n[1-a(w)]}{w} \qquad (6{\cdot}2)$$

and

$$\frac{d}{dw_\tau}\frac{[1-a(w_\tau)]}{w_\tau} = 0. \qquad (6{\cdot}3)$$

The function $a(w)$ decreases steadily from 1 to 0 as w increases from 0 to ∞, and a typical curve is shown in figure 1. The line PQ joining the points $(0, 1)$ and $(n/S, 0)$ cuts the curve at R_1, R_2 whose values of $w = w_1$, w_2 satisfy (6·2). Both points therefore give possible states of equilibrium under load S, the ordinate at each point being the fraction of the total number n of threads which have broken at equilibrium. It is possible to give a physical interpretation to the region between w_1 and w_2. When the load S is applied to the bundle, threads break until a fraction

$a(w_2)$ have given way, at which point equilibrium is established. If now additional threads are *cut*, equilibrium is still maintained until the total fraction removed is $a(w_1)$; beyond this point the cutting of an extra thread causes the whole bundle to collapse. In fact the region between w_1 and w_2 gives all possible states of equilibrium under a load S. Perhaps the argument is more clearly seen if it is noted that the ordinate of the *line* PQ gives the fraction of threads actually broken when the load on each thread is $1/w$, while the ordinate of the *curve* represents the fraction which must break under this load. Equilibrium is only possible when the former exceeds the latter.

The condition (6·3) which holds at breaking load is satisfied when PQ is tangent to the curve at w_T. When PQ is steeper than this, it does not intersect the curve at any point and there are no values of w at which equilibrium can be established.

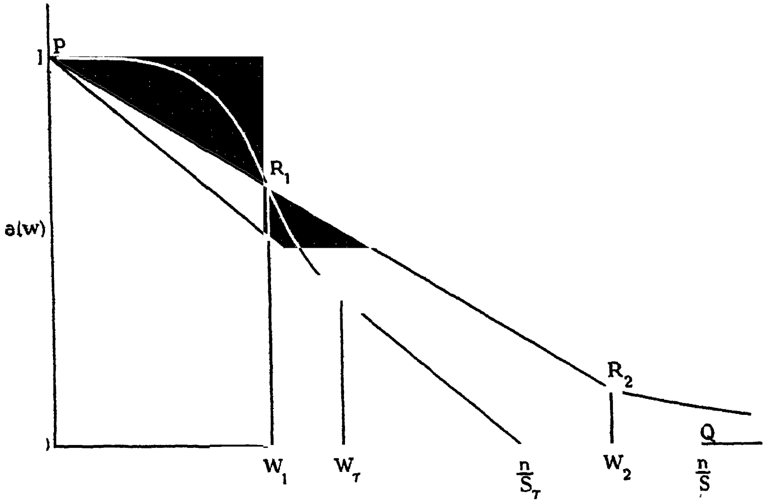


FIGURE 1

Small bundles

7. When the bundle is small, the sample of n threads may still be considered as a finite population in its own right and many of the results proved are directly applicable, but care has to be exercised in certain cases. For example, in (5·5), $\bar{\alpha}$ is the average value of α over those threads which survive at the point of rupture of the bundle.

In the simplest case when all threads have the same value of α, let the strengths of the n threads be arranged in descending order of magnitude as $\sigma_1, \sigma_2, \ldots, \sigma_n$. Then the condition for equilibrium with r survivors under a load S is

$$S \leqslant r\sigma_r \tag{7·1}$$

and the breaking load of the bundle is given by

$$S_T = \max(r\sigma_r). \tag{7·2}$$

This rather curious rule may be illustrated by an example. Suppose that six threads when tested singly snapped at loads given in the first line of table 1. Their descending order is written out in the second line and finally the product of load × order is given. Then the rule states that if all six threads had been tested as a bundle, its breaking load would have been given by the greatest of these products, that is, by 22·0.

TABLE 1

load σ	3·2	5·2	4·4	8·2	6·1	5·7
order r	6	4	5	1	2	3
$\sigma \times r$	19·2	20·8	22·0*	8·2	12·2	17·1

A graphical representation analogous to figure 1 is also available for small bundles, the curve for $a(w)$ being replaced by the survivor diagram of w_r for the sample, which descends from 1 to 0 in steps of $1/n$. The example of table 1 is shown diagrammatically in figure 2.

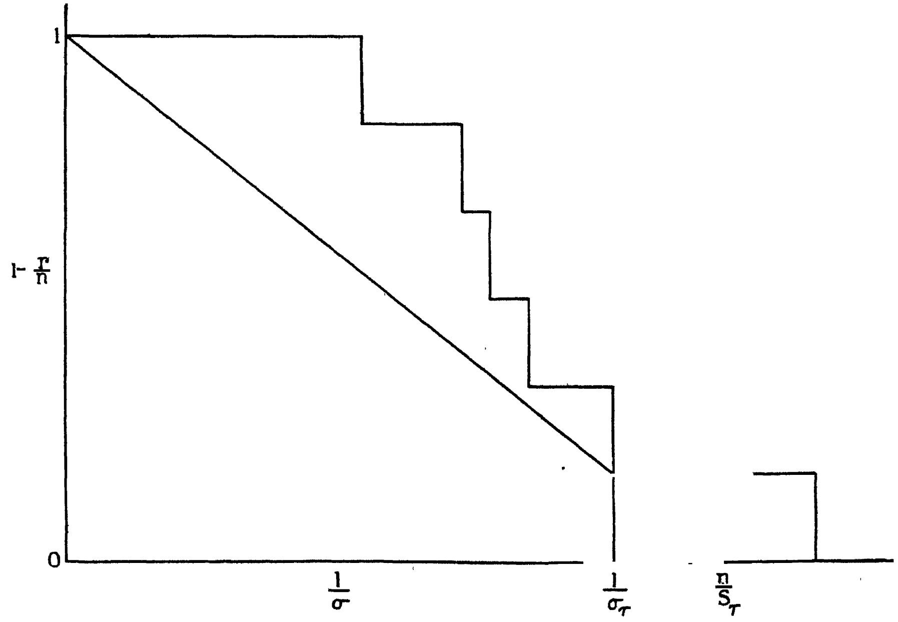


FIGURE 2

The formulae for very large bundles were obtained on the hypothesis that they all tend to have the same breaking load. It emerges later (§ 23) that while in the majority of cases this assumption is true, for certain populations of threads the breaking loads of large bundles do not tend to a limiting value with negligible dispersion. But even in such a case, the formulae may still be validly applied to the bundle considered as a finite population.

THE PROBABILITY DISTRIBUTION OF BUNDLE STRENGTH

8. From now on the discussion is confined to threads having a constant value of α. It is assumed that the bundle is randomly sampled from a population of such threads and we proceed to derive an expression for the chance that a bundle of n threads is of strength less than S.

It is evident on considering the breakage of successive threads that if the bundle succumbs under load S the conditions to be satisfied are

$$\left.\begin{aligned} 0\leqslant\sigma_n &\leqslant\frac{S}{n}, \\ \sigma_n\leqslant\sigma_{n-1}&\leqslant\frac{S}{n-1}, \\ \cdots\cdots&\cdots\cdots\cdots \\ \sigma_3\leqslant\sigma_2 &\leqslant\frac{S}{2}, \\ \sigma_2\leqslant\sigma_1 &\leqslant S, \end{aligned}\right\} \tag{8·1}$$

where $\sigma_n, \sigma_{n-1}, \ldots, \sigma_2, \sigma_1$ are the strengths of threads breaking in order $1, 2, \ldots, n-1, n$ and the chance of the event (8·1) occurring is

$$B_n = n!\int_0^{S/n}\theta(\sigma_n)\,d\sigma_n\int_{\sigma_n}^{S/(n-1)}\theta(\sigma_{n-1})\,d\sigma_{n-1}\cdots\int_{\sigma_3}^{\frac{1}{2}S}\theta(\sigma_2)\,d\sigma_2\int_{\sigma_2}^{S}\theta(\sigma_1)\,d\sigma_1$$

the factor $n!$ allowing for all possible ways of arranging the threads. Setting

$$b(\sigma)=\int_0^{\sigma}\theta(u)\,du,\quad b_r=b\left(\frac{S}{r}\right),\quad x_r=b(\sigma_{r+1})$$

the formula is more simply written as

$$B_n = n!\int_0^{b_n}dx_{n-1}\int_{x_{n-1}}^{b_{n-1}}dx_{n-2}\cdots\int_{x_2}^{b_2}dx_1\int_{x_1}^{b_1}dx_0. \tag{8·2}$$

The chance (8·2) may be expressed in a variety of ways, some of which are now given.

9. The slightly more general function

$$B_n(x) = n!\int_x^{b_n}dx_{n-1}\int_{x_{n-1}}^{b_{n-1}}dx_{n-2}\cdots\int_{x_2}^{b_2}dx_1\int_{x_1}^{b_1}dx_0 \tag{9·1}$$

satisfies the equation $$\frac{\partial B_n(x)}{\partial x}=-nB_{n-1}(x)$$

with $B_0(x)=1$ and has the property that $B_n(b_n)=0$, consequently a Taylor expansion leads to

$$\begin{aligned} 0 = B_n(x)-n(b_n-x)\,B_{n-1}(x)+\frac{n(n-1)}{2!}(b_n-x)^2\,B_{n-2}(x)-\ldots \\ +(-)^{n-1}(b_n-x)^{n-1}B_1(x)+(-)^n(b_n-x)^n. \end{aligned} \tag{9·2}$$

Solving the set of equations with $n = n, n-1, n-2, \ldots$ and $x = 0$ we obtain

$$B_n = n!\begin{vmatrix} b_n & \frac{b_n^2}{2!} & \frac{b_n^3}{3!} & \cdots & \frac{b_n^{n-1}}{(n-1)!} & \frac{b_n^n}{n!} \\ 1 & b_{n-1} & \frac{b_{n-1}^2}{2!} & \cdots & \frac{b_{n-1}^{n-2}}{(n-2)!} & \frac{b_{n-1}^{n-1}}{(n-1)!} \\ 0 & 1 & b_{n-2} & \cdots & \frac{b_{n-2}^{n-3}}{(n-3)!} & \frac{b_{n-2}^{n-2}}{(n-2)!} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 1 & b_1 \end{vmatrix}.$$

The corresponding formula for $B_n(x)$ is given on writing $b_r - x$ for b_r in this expression.

10. There are two useful ways in which B_n can be expanded in series.

(i) The first is most directly arrived at from first principles. Consider the whole range of strength from 0 to ∞ to be divided into intervals bounded by S/n, $S/n-1$, $\ldots$, $S/2$, S. The chance of a thread having strength between $S/r-1$ and S/r is $b_{r-1} - b_r$. If a bundle of n threads breaks under load S, it must at least satisfy the condition that none of its threads exceeds S in strength, otherwise the last survivor would not break. The chance that the bundle contains p_1 threads between $S/2$ and S, p_2 between $S/3$ and $S/2$, $\ldots$, p_{n-1} between $S/n-1$ and S/n and p_n less than S/n, where $p_1 + p_2 + \ldots + p_n = n$ is

$$\frac{n!(b_1-b_2)^{p_1}(b_2-b_3)^{p_2}\ldots(b_{n-1}-b_n)^{p_{n-1}}b_n^{p_n}}{p_1!\,p_2!\cdots p_{n-1}!\,p_n!}.$$

But further conditions have to be satisfied if the bundle is to break under load S. For equilibrium never to be possible, the number of threads in the bundle which are less than S/r in strength must be at least equal to r, in which case the p's have to satisfy conditions (10·11) or their equivalent form (10·12), viz.

$$\left.\begin{array}{l} p_n \geqslant 1 \\ p_n + p_{n-1} \geqslant 2 \\ p_n + p_{n-1} + p_{n-2} \geqslant 3 \\ \cdots\cdots\cdots\cdots\cdots \\ p_n + p_{n-1} + \ldots + p_2 \geqslant n-1 \\ p_n + p_{n-1} + \ldots + p_2 + p_1 = n \end{array}\right\}, \quad (10\cdot11) \qquad \left.\begin{array}{l} p_1 \leqslant 1 \\ p_1 + p_2 \leqslant 2 \\ p_1 + p_2 + p_3 \leqslant 3 \\ \cdots\cdots\cdots\cdots\cdots \\ p_1 + p_2 + \ldots + p_{n-1} \leqslant n-1 \\ p_1 + p_2 + \ldots + p_n = n \end{array}\right\}. \quad (10\cdot12)$$

The chance of the bundle breaking under load S is therefore given by

$$B_n = \sum_p \frac{n!(b_1-b_2)^{p_1}(b_2-b_3)^{p_2}\ldots(b_{n-1}-b_n)^{p_{n-1}}b_n^{p_n}}{p_1!\,p_2!\cdots p_{n-1}!\,p_n!}, \qquad (10\cdot2)$$

where the p's are summed over all values consistent with (10·11) or (10·12). The corresponding formula for $B_n(x)$ is obtained on writing $b_r - x$ for b_r.

(ii) The second series expansion is the multiple Taylor expansion of the determinant in powers of the b_r's. We have

$$B_n = (B_n)_0 + \sum_r b_r \left(\frac{\partial B_n}{\partial b_r}\right)_0 + \frac{1}{2!} \sum_{r>s}\sum b_r b_s \left(\frac{\partial^2 B_n}{\partial b_r \partial b_s}\right)_0 + \dots$$

where $(\ )_0$ means that all the b_r's are replaced by 0 in the bracketed expression. Many of the terms of this expansion vanish on account of two rows of the determinant becoming identical after differentiation when the b_r's are made zero, and in fact the only non-vanishing terms are obtained as follows. Numbering the rows from the bottom upwards, first the r_1th row is differentiated r_1 times, then the (r_1+r_2)th row r_2 times, the $(r_1+r_2+r_3)$th row r_3 times and so on, till finally the nth (top) row is differentiated r_m times where $r_1+r_2+\dots+r_m = n$, it being essential to differentiate the top row at least once. If the resulting determinant is then rearranged with the r_1th row at the bottom, the (r_1+r_2)th row in place of the r_1th, the $(r_1+r_2+r_3)$th in place of the (r_1+r_2)th and so on, the sign of the determinant is changed by

$$(-)^{r_1-1+r_2-1+\dots+r_n-1} = (-)^{n-m}$$

and its value becomes unity. The expansion is therefore

$$B_n = \sum_{m=1}^{n} \sum_r (-)^{n-m} \frac{n!\, b_{r_1}^{r_1} b_{r_1+r_2}^{r_2} \dots b_{r_1+r_2+\dots+r_{m-1}}^{r_{m-1}} b_n^{r_m}}{r_1!\, r_2! \dots r_{m-1}!\, r_m!} \tag{10·3}$$

summed over all values of the r's such that $r_t \geqslant 1$, all t, and $r_1+r_2+\dots+r_m = n$.

Using this expansion, the following formulae are obtained

$$B_0 = 1, \quad B_1 = b_1, \quad B_2 = 2b_1b_2 - b_1^2,$$
$$B_3 = 6b_1b_2b_3 - 3b_1^2b_3 - 3b_1b_3^2 - b_3^2,$$
$$B_4 = 24b_1b_2b_3b_4 - 12b_1^2b_3b_4 - 12b_1b_3^2b_4 - 12b_1b_2b_4^2 - 4b_3^3b_4 + 6b_1^2b_4^2 - 4b_1b_4^3 - b_4^4$$

though in practice it is perhaps simpler to build them up by means of the recurrence formula (9·2) with $x = 0$. For higher values of n the formulae rapidly become unmanageable, and it is necessary to seek an asymptotic expression for B_n. Before considering this question, however, mention must be made of two forms of $b(s)$ which, though not common ones in practice, give rise to specially simple B_n expressions for all values of n.

11. Consider the case where the threads have only two possible strengths ρ and $\nu\rho$, that is

$$\left.\begin{aligned} b(s) &= 0 \quad (0 \leqslant s < \rho), \\ b(s) &= b \quad (\rho \leqslant s < \nu\rho), \\ b(s) &= 1 \quad (\nu\rho \leqslant s), \end{aligned}\right\} \tag{11·1}$$

and let a load S be applied to a bundle of n of such threads.

We observe first that $b_n = 0$ and hence $B_n = 0$ whenever $S < n\rho$. Divide up the range of S into the intervals $r\nu\rho \leqslant S < (r+1)\nu\rho$. Then for values of $S \geqslant n\rho$ we have, when S is in the rth interval,

$$b_m = b\left(\frac{S}{m}\right) = 1 \quad (m \leqslant r),$$

$$b_m = b \quad (m > r).$$

This holds for all intervals having $r > \left[\frac{n}{\nu}\right]$ and also for that part of the interval $r = \left[\frac{n}{\nu}\right]$ in which $S \geqslant n\rho$.

Consider now the series expansion (10·2). When $S \geqslant n\rho$ and S is in the rth interval the only non-vanishing terms are those in which $p_m = 0$ except for $m = r$ and $m = n$. Writing $p_r = p$, $p_n = n - p$ conditions (11·12) imply that $p \leqslant r$ and the expression for B_n becomes, with this special form for $b(s)$,

$$\left.\begin{aligned} B_n &= 0 && \text{when} \quad S < n\rho, \\ B_n &= \sum_{p=0}^{r} \frac{n!(1-b)^p b^{n-p}}{p!(n-p)!} && \text{when} \quad r\nu\rho \leqslant S < (r+1)\nu\rho \quad \text{and} \quad S \geqslant n\rho. \end{aligned}\right\} \tag{11·2}$$

The strength of the bundle can therefore only have values $r\nu\rho$ $\left(r > \left[\frac{n}{\nu}\right]\right)$ or $n\rho$, the probability of the value $r\nu\rho$ being the coefficient of t^r in $[b + t(1-b)]^n$ and the probability of $n\rho$ the sum of the coefficients of t^p for values of $p \leqslant \left[\frac{n}{\nu}\right]$. The approximate form of B_n when n is large is in this case easily deduced from the normal approximation to the binomial distribution. For very large n there are two distinct limiting forms according to whether the 'collapsed' part contains the greater or lesser number of binomial terms, the appropriate conditions being $\nu(1-b) < 1$ or > 1 respectively. In the first case all values concentrate at $S = n\rho$; in the second they concentrate at $n\nu\rho(1-b)$. These ultimate limiting forms are immediately obvious on applying the method of § 6 for very large bundles.

12. The second form of $b(s)$ is that in which the weakness $w = 1/s$ of a thread has an equal chance of lying anywhere in the interval $0 \leqslant w \leqslant 1/s_0$ but cannot exceed $1/s_0$, that is

$$\left.\begin{aligned} b(s) &= 0 && (0 \leqslant s < s_0) \\ b(s) &= 1 - \frac{s_0}{s} && (s_0 \leqslant s). \end{aligned}\right\} \tag{12·1}$$

We now show that the chance that a bundle of n threads sampled from such a population has strength less than S is

$$\left.\begin{aligned} B_n &= 0 && (0 \leqslant S < ns_0) \\ B_n &= 1 - \frac{ns_0}{S} && (ns_0 \leqslant S) \end{aligned}\right\} \tag{12·2}$$

so that for all values of n the probability distribution of S/n is simply $b(S/n)$.

The proof is as follows. Evidently $b_n = 0$ when $0 \leqslant S < ns_0$ and hence B_n vanishes for values of S in this interval. Let $S \geqslant ns_0$ and insert $b_r = 1 - \frac{rs_0}{S}$ in (9·1). The function

$$B_n(x) = (1-x)^n - \frac{ns_0}{S}(1-x)^{n-1}$$

is such that $B_0(x) = 1$, $B_n\left(1 - \frac{ns_0}{S}\right) = 0$ and it satisfies

$$\frac{\partial B_n(x)}{\partial x} = -nB_{n-1}(x).$$

It must therefore be identical with (9·1) for the particular form of b_r chosen, and putting $x = 0$ the result follows.

The method already given for large bundles breaks down in this case since

$$s[1-b(s)] = s_0$$

for $s_0 \leqslant s$ and there is no unique value of s_r which maximizes it. It will be of interest later to recall (12·2) in connexion with the asymptotic form of B_n when $b(s)$ behaves like $1 - s_0/s$ for large s (§ 23).

THE ASYMPTOTIC BEHAVIOUR OF B_n FOR LARGE n

13. It has been shown that as n becomes very large the bundle strength is likely to concentrate near a value $S_r = n \max s[1-b(s)]$. No indication has so far been given, however, except in two special cases of the form assumed by B_n in the vicinity of S_r, and the remainder of the paper is devoted to a study of the asymptotic behaviour of B_n for general forms of the parent distribution $b(s)$. The mode of attack adopted here provides the dominant term of the asymptotic expansion together with the order of magnitude of the approximation involved.

It is first necessary to derive an important identity upon which the subsequent theory depends. Let us introduce the notation

$$(n, m) = \int_0^{b_n} dx_{n-1} \int_{x_{n-1}}^{b_{n-1}} dx_{n-2} \cdots \int_{x_{m+1}}^{b_{m+1}} dx_m$$

$$= \begin{vmatrix} b_n & \frac{b_n^2}{2!} & \cdots & \frac{b_n^{n-m-1}}{(n-m-1)!} & \frac{b_n^{n-m}}{(n-m)!} \\ 1 & b_{n-1} & \cdots & \frac{b_{n-1}^{n-m-2}}{(n-m-2)!} & \frac{b_{n-1}^{n-m-1}}{(n-m-1)!} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 1 & b_{m+1} \end{vmatrix} \qquad (13·1)$$

and $(n, n) = 1$, so that, for example, $B_n = n!(n, 0)$. Now in the determinant $(n, 0)$ multiply the $(n-1)$th column by $-x$, the $(n-2)$th column by $\frac{x^2}{2!}$, ..., the 1st column by $(-)^{n-1}\frac{x^{n-1}}{(n-1)!}$ and add them to the last column, thus obtaining

$$(n,0) = \left|\begin{array}{cccccc} b_n & \frac{b_n^2}{2!} & \frac{b_n^3}{3!} & \cdots & \frac{b_n^{n-1}}{(n-1)!} & \frac{(b_n-x)^n}{n!}-(-)^n\frac{x^n}{n!} \\ 1 & b_{n-1} & \frac{b_{n-1}^2}{2!} & \cdots & \frac{b_{n-1}^{n-2}}{(n-2)!} & \frac{(b_{n-1}-x)^{n-1}}{(n-1)!} \\ 0 & 1 & b_{n-2} & \cdots & \frac{b_{n-2}^{n-3}}{(n-3)!} & \frac{(b_{n-2}-x)^{n-2}}{(n-2)!} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 1 & b_1 - x \end{array}\right|.$$

The expansion of the determinant in terms of the last column may then be rearranged to give

$$\frac{x^n}{n!} \equiv \sum_{m=0}^{n} (n,m)\frac{(x-b_m)^m}{m!} \quad \text{for all } x, \tag{13·21}$$

or, more generally, after differentiating r times,

$$\frac{x^{n-r}}{(n-r)!} \equiv \sum_{m=r}^{n} (n,m)\frac{(x-b_m)^{m-r}}{(m-r)!} \quad \text{for all } x \text{ and } r \geqslant 0. \tag{13·22}$$

One observes in passing that (n, m) is related to the chance $B_{n,m}$ that m threads will survive under a load S by the formula

$$B_{n,m} = \frac{n!}{m!}(n,m)(1-b_m)^m.$$

Substituting in (13·21) leads to

$$x^n = \sum_{m=0}^{n} B_{n,m}\left(\frac{x-b_m}{1-b_m}\right)^m, \quad \text{all } x.$$

For example, when $x = 1$ the result $1 = \sum_{m=0}^{n} B_{n,m}$ simply expresses the fact that values of m from 0 to n exhaust all possible contingencies.

The identities (13·21) and (13·22) are not in the form most convenient for our purpose. Writing $x = 1/\lambda$, (13·21) becomes

$$1 \equiv \sum_{m=0}^{n} \frac{n!(n,m)}{m!}\lambda^{n-m}(1-\lambda b_m)^m, \quad \text{all } \lambda,$$

and by analogy with the binomial distribution, one is led to consider the expression

$$Q_{n,m} = \frac{(n-m)!(n,m)}{b_m^{n-m}} = (n-m)!\int_0^{b_n/b_m} dx_{n-1}\int_{x_{n-1}}^{b_{n-1}/b_m} dx_{n-2}\cdots\int_{x_{m+1}}^{b_{m+1}/b_m} dx_m, \tag{13·3}$$

which satisfies, and is in fact uniquely determined by,

$$1 = \sum_{m=0}^{n} Q_{n,m} \frac{n!(\lambda b_m)^{n-m}(1-\lambda b_m)^m}{m!(n-m)!}, \quad \text{all } \lambda, \tag{13·4}$$

or the more general identity

$$1 = \sum_{m=r}^{n} Q_{n,m} \frac{(n-r)!\,(\kappa b_m)^{n-m}(1-\kappa b_m)^{m-r}}{(m-r)!\,(n-m)!}, \quad \text{all } \kappa, \text{ all } r \geqslant 0, \tag{13·5}$$

where κ is written in place of λ for a reason that will be apparent later. Note that $Q_{n,0}$ is simply B_n.

The function $T_{n,m,r}$

14. When n is large, (13·4) and (13·5) approximate to integral equations and the behaviour of $Q_{n,m}$ for large n is evidently related to that of

$$T_{n,m,r} = \frac{(n-r)!\,(\kappa b_m)^{n-m}(1-\kappa b_m)^{m-r}}{(m-r)!\,(n-m)!}.$$

A substantial digression is therefore made at this stage to investigate the asymptotic behaviour of $T_{n,m,r}$.

It will be found simplest first to discuss the special case

$$T_{n,m} = \frac{n!(\lambda b_m)^{n-m}(1-\lambda b_m)^{m-r}}{m!(n-m)!},$$

and to extend the results afterwards to the more general function. The important range of λ is $0 \leqslant \lambda \leqslant 1$ in which $T_{n,m}$ is never negative. In the case of the ordinary binomial distribution $\dfrac{n!\,p^{n-m}(1-p)^m}{m!(n-m)!}$ there are two distinct limiting forms for large n. When neither p nor $1-p$ is small the binomial is approximated to by a continuous normal distribution having the same first and second moments np and $np(1-p)$ respectively. On the other hand, when p is small the appropriate limiting form is the discrete Poisson distribution with parameter np, and similarly for small $1-p$. In the present more general case it is useful to preserve the distinction, and we first discuss various continuous limiting forms corresponding to the normal approximation.

Continuous limiting forms of $T_{n,m}$

15. Assuming that both m and $n-m$ are $O(n)$ so that neither m/n nor $1-m/n$ is small, Stirling's formula gives

$$T_{n,m} = \frac{n^{n+\frac{1}{2}}(\lambda b_m)^{n-m}(1-\lambda b_n)^m}{\sqrt{2\pi}\, m^{m+\frac{1}{2}}(n-m)^{n-m+\frac{1}{2}}} \quad [1+O(n^{-1})].$$

Writing $m/n = z$, $1/n = dz$, $S = -n\zeta$, $T_{n,m} = T_n(z)\,dz$ and $a(1/s) = b(s)$ as defined in §7, the approximation is

$$T_n(z)\,dz = \sqrt{\left(\frac{n}{2\pi}\right)} \frac{dz}{\sqrt{[z(1-z)]}} \left[\frac{\lambda a(z/\zeta)}{1-z}\right]^{n(1-z)} \left[\frac{1-\lambda a(z/\zeta)}{z}\right]^{nz} [1+O(n^{-1})]. \tag{15·1}$$

Clearly the important regions for $T_n(z)$ are near the maxima of

$$F(z) = \left[\frac{\lambda a(z/\zeta)}{1-z}\right]^{(1-z)} \left[\frac{1-\lambda a(z/\zeta)}{z}\right]^{z}$$
$$= \left[1-\frac{(1-z-\lambda a)}{1-z}\right]^{(1-z)} \left[1+\frac{(1-z-\lambda a)}{z}\right]^{z}$$
$$= \exp\left\{-\tfrac{1}{2}(1-z-\lambda a)^2\left(\frac{1}{1-z}+\frac{1}{z}\right)-\tfrac{1}{3}(1-z-\lambda a)^3\left(\frac{1}{(1-z)^2}-\frac{1}{z^2}\right)-\dots\right\}. \quad (15{\cdot}2)$$

Differentiating, $\dfrac{F'}{F} = \dfrac{(1-z\lambda a)}{z(1-z)}\left(1+\dfrac{\lambda}{\zeta}a'\right)$ + higher powers of $(1-z-\lambda a)$,

$$\frac{F''}{F}-\left(\frac{F'}{F}\right)^2 = -\frac{\left(1+\dfrac{\lambda}{\zeta}a'\right)}{z(1-z)} + \text{terms containing } (1-z-\lambda a).$$

The condition $F'(z) = 0$ is thus at least satisfied when

$$1-z-\lambda a(z/\zeta) = 0 \quad (15{\cdot}3)$$

and at the roots $z = z_r$ of this equation it is found that

$$F(z_r) = 1, \quad F''(z_r) = -\left[1+\frac{\lambda}{\zeta}a'(z_r/\zeta)\right]^2 \Big/ z_r(1-z_r). \quad (15{\cdot}4)$$

It follows that the roots z_r are all maxima of $F(z)$.

Moreover, at all other points in $0<z<1$ one must necessarily have $F(z)<1$. For consider $F = e^G$ as a function of λ. Turning points occur when $G_\lambda = 0$, where

$$G_\lambda = a\left\{\frac{1-z}{\lambda a}-\frac{z}{1-\lambda a}\right\} = \frac{(1-z-\lambda a)}{\lambda(1-\lambda a)}$$

and for given z there is only one value of λ satisfying $G_\lambda = 0$, namely, that making $1-z-\lambda a = 0$. Furthermore

$$G_{\lambda\lambda} = -a^2\left\{\frac{1-z}{(\lambda a)^2}+\frac{z}{(1-\lambda a)^2}\right\},$$

showing that $G_{\lambda\lambda}<0$ for all z and λ lying between 0 and 1. Hence G has a unique maximum at the point $\lambda = (1-z)/a$ and so has F for the same λ, taking the value $F = 1$ there, and decreasing steadily on either side of it. No values of z and λ between 0 and 1 can therefore be found at which $F \geqslant 1$ and $1-z-\lambda a \neq 0$, for if such a value existed, then at $\lambda = (1-z)/a$ one should also have $F = 1$, $1-z-\lambda a = 0$ which would imply $G_{\lambda\lambda} \geqslant 0$ at some point of the range.

Considering F once more as a function of z for fixed λ, we conclude that when n is large, F^n is negligible except near the roots $z = z_r$ of (15·3) at which it approximates to 1, and consequently $T_n(z)$ consists of a series of isolated peaks situated at these roots.

The positions of the roots are conveniently examined by means of a diagram (figure 3) similar to that introduced in § 7. A typical curve for $a(w)$ is shown. The line joining the point $(0, 1/\lambda)$ to the point $(1/\zeta, 0)$ intersects the curve at points

z_1/ζ, z_2/ζ, z_3/ζ, and it is evident that z_1, z_2, z_3 so found are the only positive roots of $1-z-\lambda a(z/\zeta) = 0$. With an $a(w)$ curve of the form shown a maximum of 3 positive real roots is possible with suitable values of ζ and λ, and in general the maximum number of roots is odd.

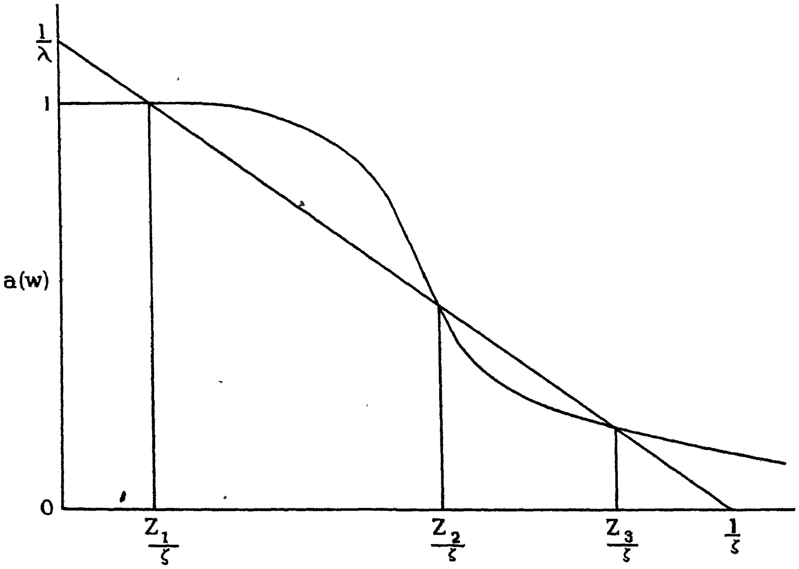


FIGURE 3

16. Consider first the behaviour of $T_n(z)$ near a simple real root z_r such that neither z_r nor $1-z_r$ is small. On substituting

$$1-z-\lambda a(z/\zeta) = -\left[1+\frac{\lambda}{\zeta}a'(z_r/\zeta)\right]^2 (z-z_r)^2 + O(z-z_r)^3$$

in (15·2) the appropriate expansion of (15·1) is found to be

$$T_n(z) = \sqrt{\left(\frac{n}{2\pi z_r(1-z_r)}\right)} \exp\left[-\frac{n\left[1+\frac{\lambda}{\zeta}a'(z_r/\zeta)\right]^2 (z-z_r)^2}{2z_r(1-z_r)}\right]\{1+O(n(z-z_r)^3)+O(z-z_r)\}.$$

Since the exponential is negligibly small except when $n(z-z_r)^2 = O(1)$, only the range $z-z_r = O(n^{-\frac{1}{2}})$ need be considered and the approximation factor is $1+O(n^{-\frac{1}{2}})$. It will be found more useful, however, to write the result in the form

$$T_n(z) = \sqrt{\left(\frac{n}{2\pi z_r(1-z_r)}\right)} \exp\left[-\frac{n\left[1+\frac{\lambda}{\zeta}a'(z_r/\zeta)\right]^2 (z-z_r)^2}{2z_r(1-z_r)}\right]$$
$$\times\{1+A(z-z_r)+Bn(z-z_r)^3+O(n^{-1})\}, \quad (16{\cdot}1)$$

where the constants A and B are $O(1)$.

17. Suppose next that the two roots z_2 and z_3, which may be real or complex, are such that $|z_2 - z_3|$ is small. It is assumed that neither $|z_r|$ nor $|1 - z_r|$ is small $(r = 2, 3)$.

Write
$$\epsilon = \tfrac{1}{2}(z_3 - z_2)$$
and
$$f(z) = 1 - z - \lambda a(z/\zeta).$$
Let $\bar{z}$ be that value of z near z_2, z_3 which makes $f'(z) = 0$; then since
$$f(z) = f(\bar{z}) + \tfrac{1}{2}(z - \bar{z})^2 f''(\bar{z}) + O(z - \bar{z})^3$$
and
$$f(z_2) = f(z_3) = 0, \quad f''(\bar{z}) = -\frac{\lambda}{\zeta^2} a''(\bar{z}/\zeta)$$
it follows, if $f''(\bar{z})$ is not small and $z - \bar{z} = O(\epsilon)$, that
$$f(z) = -\frac{\lambda a''(\bar{z}/\zeta)}{2\zeta^2}[(z - \bar{z})^2 - \epsilon^2] + O(\epsilon^3) \tag{17·1}$$
and
$$\bar{z} = \tfrac{1}{2}(z_2 + z_3) + O(\epsilon^2).$$
The sign of ϵ^2 may be either positive or negative according as the roots are real or complex.

If the parameters λ and ζ are altered until $z_2 = z_3$, we shall call the coincident value of the roots a *critical point* of z. In particular, two critical points z_τ, z_c are obtained by (i) keeping λ fixed and changing ζ or (ii) keeping ζ fixed and changing λ, respectively, the corresponding critical values of ζ and λ being ζ_τ and λ_c. It will be shown that both z_τ and z_c differ from $\bar{z}$ by $O(\epsilon^2)$. They are illustrated graphically in figure 4.

Expressions for $T_n(z)$ are now obtained in terms of z_τ and z_c respectively.

(i) Put $w = z/\zeta$, $w_\tau = z_\tau/\zeta_\tau$, and let
$$\zeta(w) = \frac{1 - \lambda a(w)}{w} = \zeta_\tau + (w - w_\tau)\,\zeta'_\tau + \tfrac{1}{2}(w - w_\tau)^2\,\zeta''_\tau + O(w - w_\tau)^3,$$
where the suffix τ denotes evaluation at w_τ. By (15·3),
$$\zeta(w_2) = \zeta(w_3) = \zeta, \quad \zeta'_\tau = 0, \quad \zeta''_\tau = -\frac{\lambda a''_\tau}{w_\tau}.$$
In general ζ''_τ is not small, so that
$$w_\tau = \tfrac{1}{2}(w_2 + w_3) + O(\epsilon^2), \quad \epsilon^2 = \tfrac{1}{4}\zeta_\tau^2(w_2 - w_3)^2 + O(\epsilon^3)$$
or, to the same order
$$z_\tau = \bar{z} + O(\epsilon^2), \quad \epsilon^2 = -\frac{2z_\tau\zeta_\tau}{\lambda a''_\tau}(\zeta - \zeta_\tau) + O(\epsilon^2),$$
and inserting these results in (16·1) and (15·2) the form of $F(z)$ in the range $z - z_\tau = O(\epsilon)$ is found to be
$$F(z) = \exp\left\{-\frac{1}{2z_\tau(1 - z_\tau)}\left[\frac{\lambda a''_\tau}{2\zeta_\tau^2}(z - z_\tau)^2 + \frac{z_\tau}{\zeta_\tau}(\zeta - \zeta_\tau)\right]^2 + O(\epsilon^5)\right\}.$$

Now ϵ has to be at most $O(n^{-\frac{1}{4}})$ for otherwise $T_n(z)$ in (15·1) becomes small at $z - z_\tau$ and z_2, z_3 may then be considered as isolated simple roots. Also, in the vicinity of z_τ, $T_n(z)$ is negligible outside a range $z-z_\tau = O(n^{-\frac{1}{4}})$ and its approximate form is therefore

$$T_n(z) = \sqrt{\left(\frac{n}{2\pi z_\tau(1-z_\tau)}\right)}\exp\left\{-\frac{n}{2z_\tau(1-z_\tau)}\left[\frac{\lambda a''_\tau}{2\zeta^2_\tau}(z-z_\tau)^2+\frac{z_\tau}{\zeta_\tau}(\zeta-\zeta_\tau)\right]^2\right\}\{1+O(n^{-\frac{1}{4}})\} \tag{17·2}$$

provided $$z_\tau = O(1), \quad 1-z_\tau = O(1).$$

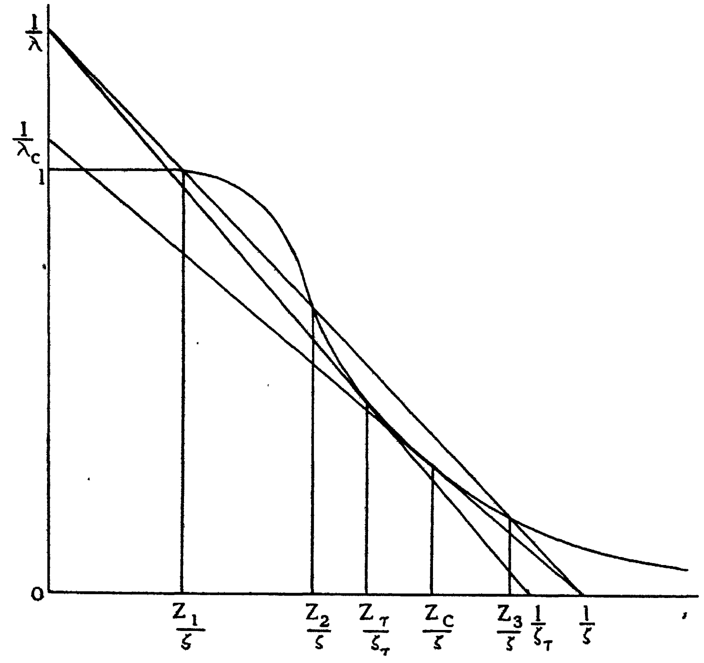


FIGURE 4

(ii) Let $$\lambda(z) = \frac{1-z}{a(z/\zeta)} = \lambda_c+(z-z_c)\lambda'_c+\tfrac{1}{2}(z-z_c)^2\lambda''_c+O(z-z_c)^3.$$

By definition $\lambda(z_2) = \lambda(z_3) = 0$, and $\lambda'_c = 0$, $\lambda''_c = -\lambda_c a''_c/\zeta^2 a_c$, the suffix c denoting evaluation at z_c. Since λ''_c is not usually small, it follows that

$$z_c = \bar{z}+O(\epsilon^2), \quad \epsilon^2 = -\frac{2\zeta^2 a_c}{\lambda_c a''_c}(\lambda-\lambda_c)+O(\epsilon^3)$$

and, as before, the approximate form of $T_n(z)$ is

$$T_n(z) = \sqrt{\left(\frac{n}{2\pi z_c(1-z_c)}\right)}\exp\left\{-\frac{n}{2z_c(1-z_c)}\left[\frac{\lambda_c a''_c}{2\zeta^2}(z-z_c)^2+\frac{a_c}{\zeta}(\lambda-\lambda_c)\right]^2\right\}\{1+O(n^{-\frac{1}{4}})\}. \tag{17·3}$$

There is one other continuous limiting form of $T_n(z)$ which has to be considered. When the first root z_1 is isolated and small the approximation to $T_n(z)$ in its vicinity is no longer given by (16·1), the discrete limiting form discussed below (§ 20) taking its place. But when z_1 and z_2 are close together and both small, a continuous approximation to $T_n(z)$ is still appropriate. It is best to discuss this case later (§ 19) in its more general form, after the results so far obtained have been extended to $T_{n,m,r}$.

Continuous limiting forms of $T_{n,m,r}$

18. The asymptotic behaviour of the function

$$T_{n,m,r} = \frac{(n-r)!\,(\kappa b_m)^{n-m}\,(1-\kappa b_m)^m}{(n-m)!\,(m-r)!}$$

is discussed on similar lines and the proofs will be only briefly sketched. For $T_{n,m,r}$ to be never negative in the range $r \leqslant m \leqslant n$, κ has to satisfy the condition

$$0 \leqslant \kappa \leqslant 1/b_r. \tag{18·1}$$

The continuous approximations hold when neither $(m-r)/n$ nor $1-m/n$ is small, and Stirling's formula then gives

$$T_{n,m,r} = \frac{(n-r)^{n-r+\frac{1}{2}}\,(\kappa b_m)^{n-m}\,(1-\kappa b_m)^{m-r}}{\sqrt{2\pi}\,(m-r)^{m-r+\frac{1}{2}}\,(n-m)^{n-m+\frac{1}{2}}}\,[1+O(n^{-1})].$$

Adopting the previous notation and putting in addition

$$r = nx, \quad T_{n,m,r} = T_n(z,x)\,dz$$

this becomes

$$T_n(z,x) = \sqrt{\left(\frac{n(1-x)}{2\pi(z-x)\,(1-z)}\right)}\,(1-x)^{n(1-x)}\left[\frac{\kappa a(z/\zeta)}{1-z}\right]^{n(1-z)} \times\left[\frac{1-\kappa a(z/\zeta)}{z-x}\right]^{n(z-x)}[1+O(n^{-1})], \tag{18·2}$$

where $x<z<1$, and we may write

$$\begin{aligned} F(z,x) &= (1-x)^{(1-x)}\left[\frac{\kappa a(z/\zeta)}{1-z}\right]^{(1-z)}\left[\frac{1-\kappa a(z/\zeta)}{z-x}\right]^{(z-x)} \\ &= \exp\left\{-\tfrac{1}{2}(1-z-\lambda a)^2\left(\frac{1}{1-z}+\frac{1}{z-x}\right)-\tfrac{1}{3}(1-z-\lambda a)^2\left(\frac{1}{(1-z)^2}-\frac{1}{(z-x)^2}\right)-\dots\right\}, \end{aligned} \tag{18·3}$$

where

$$\lambda = (1-x)\,\kappa. \tag{18·4}$$

By (18·1), λ must satisfy the condition

$$0 \leqslant \lambda \leqslant \lambda_x = (1-x)/a(x/\zeta). \tag{18·5}$$

$F(z,x)$ then takes its maximum value $F = 1$ at the roots z_r of $1-z-\lambda a(z/\zeta) = 0$, and the argument proceeds as before. Near an isolated simple root z_r such that $x<z_r<1$, $z_r-x = O(1)$, $1-z_r = O(1)$ the limiting form of $T_n(z,x)$ is

$$T_n(z,x) = \surd\left(\frac{n(1-x)}{2\pi(z_r-x)(1-z_r)}\right)$$
$$\times\exp\left(-\frac{n(1-x)\left[1+\frac{\lambda}{\zeta}a'(z_r/\zeta)\right]^2(z-z_r)^2}{2(z_r-x)(1-z_r)}\right)[1+A(z-z_r)+Bn(z-z_r)^3+O(n^{-1})], \tag{18·6}$$

and near a critical point the formulae corresponding to (17·2) and (17·3) are

$$T_n(z,x) = \surd\left(\frac{n(1-x)}{2\pi(z_\tau-x)(1-z_\tau)}\right)$$
$$\times\exp\left\{-\frac{n(1-x)}{2(z_\tau-x)(1-z_\tau)}\left[\frac{\lambda a''_\tau}{2\zeta_\tau^2}(z-z_\tau)^2+\frac{z_\tau}{\zeta_\tau}(\zeta-\zeta_\tau)\right]^2\right\}[1+O(n^{-\frac{1}{4}})] \tag{18·7}$$

$$T_n(z,x) = \surd\left(\frac{n(1-x)}{2\pi(z_c-x)(1-z_c)}\right)$$
$$\times\exp\left\{-\frac{n(1-x)}{2(z_c-x)(1-z_c)}\left[\frac{\lambda_c a''_c}{2\zeta^2}(z-z_c)^2+\frac{a_c}{\zeta}(\lambda-\lambda_c)\right]^2\right\}[1+O(n^{-\frac{1}{4}})]. \tag{18·8}$$

19. Suppose now that x is in the neighbourhood of the critical point. Let $z_c-x = O(\epsilon)$; then, on substituting (17·1) in (18·3), $F(z,x)$ assumes the form

$$F(z,x) = \exp\left\{-\frac{1}{2(z-x)}\left[\frac{\lambda_c a''_c}{2\zeta^2}(z-z_c)^2+\frac{a_c}{\zeta}(\lambda-\lambda_c)\right]^2+O(\epsilon^4)\right\}$$

since $1-z = 1-x+O(\epsilon)$, and (18·2) becomes

$$T_n(z,x) = \surd\left(\frac{n}{2\pi(z-x)}\right)$$
$$\times\exp\left\{-\frac{n}{2(z-x)}\left[\frac{\lambda_c a''_c}{2\zeta^2}(z-z_c)^2+\frac{a_c}{\zeta}(\lambda-\lambda_c)\right]^2\right\}[1+O(\epsilon)+O(n\epsilon^4)]. \tag{19·1}$$

For the exponent to be $O(1)$ at $z = z_c$ we must have $\epsilon = O(n^{-\frac{1}{3}})$ and so, over a range $z-z_c = O(n^{-\frac{1}{3}})$, (19·1) holds to $O(n^{-\frac{1}{3}})$.

In the discussion of the limiting form of $Q_{n,m}$, however, it will be necessary to assume $\epsilon = O(n^{-\frac{1}{4}})$ in this case. At $z = z_c$ the exponent is then $O(n^{\frac{1}{4}})$ and $T_n(z,x)$ splits into two isolated peaks at z_2, z_3. The behaviour of $T_n(z,x)$ near z_3 is of interest, and one would expect the approximation (18·6) to hold there, but to a lower order of accuracy. To study $T_n(z,x)$ near z_3, write (17·1) as

$$f(z) = -\frac{\lambda_c a''_c}{2\zeta^2}(z-z_2)(z-z_3)+O(\epsilon^3).$$

$T_n(z, x)$ is not negligible in a range of $z-z_3$ satisfying

$$\frac{n\lambda_c^2 a_c''^2}{8\zeta_c^2}\frac{(z-z_2)^2(z-z_3)^2}{(z-x)} = O(1)$$

and since $z-z_2$ and $z-x$ are $O(n^{-\frac{1}{4}})$, this implies

$$z-z_3 = O(n^{-\frac{3}{8}}).$$

The approximation near z_3 when $\epsilon = O(n^{-\frac{1}{4}})$ is then found to be

$$T_n(z, x) = \sqrt{\left(\frac{n}{2\pi(z_3-x)}\right)}$$
$$\times \exp\left[-\frac{n\lambda_c^2 a_c''^2}{2\zeta^4}\frac{(z_3-z_c)^2(z-z_3)^2}{(z_3-x)}\right][1+An^{\frac{1}{4}}(z-z_3)+Bn(z-z_3)^2+O(n^{-\frac{1}{4}})], \tag{19·2}$$

where A, B are $O(1)$ and independent of z. The formula agrees to $O(n^{-\frac{1}{4}})$ with (18·6).

Discrete limiting forms of $T_{n,m}$

20. Reverting once more to

$$T_{n,m} = \frac{n!}{m!(n-m)!}(\lambda b_m)^{n-m}(1-\lambda b_m)^m$$

the analogue of the Poisson binomial limit, when n is large but m is $O(1)$ is next discussed. It is assumed that $a(z/\zeta)$ admits of an expansion in integral powers of z near $z = 0$. The possibility of an expansion in non-integral powers is realized, and, in fact, such an expansion is to be expected in some cases in view of the limiting form (5·7). But the more general analysis is not essentially different, and for simplicity only integral powers are considered. The expansion of $a(z/\zeta)$ is then

$$a(z/\zeta) = 1+a'(0)\frac{z}{\zeta}+O(z^2),$$

i.e.

$$b_m = 1-\frac{\zeta_0 m}{\zeta\ n}+O(n^{-2}) \tag{20·1}$$

where $1/\zeta_0$ is the intercept on the horizontal axis of the tangent to the $a(w)$ curve at the origin. Two types of $a(w)$ curve are shown in figures 5 and 6. Since the important values of λ turn out to be near $\lambda = 1$, we put

$$\lambda = 1-\frac{\mu}{n}$$

where μ is $O(1)$. Then

$$(n-m)\log\lambda b_m = -(\mu+m\zeta_0/\zeta)+O(n^{-1})$$

and

$$(1-\lambda b_m) = \frac{1}{n}(\mu+m\zeta_0/\zeta)+O(n^{-2}).$$

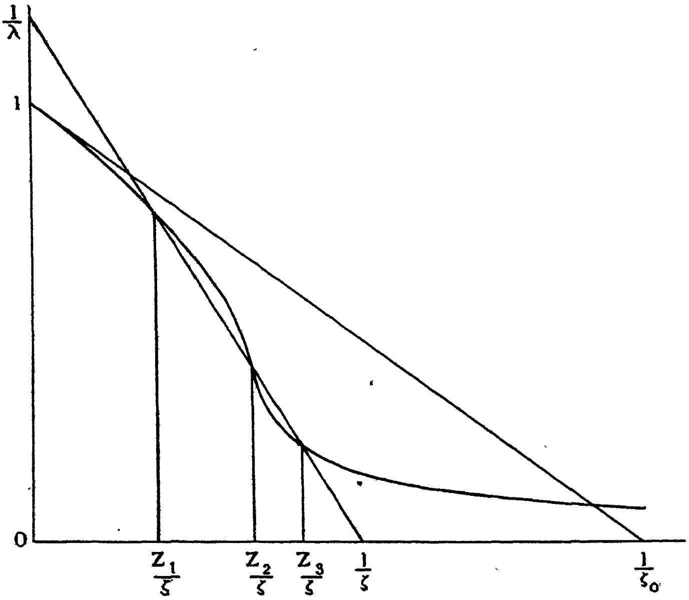


FIGURE 5

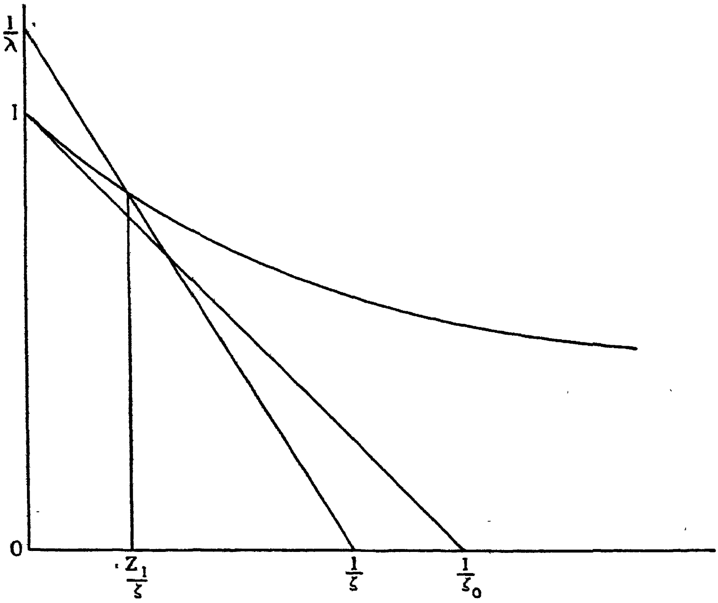


FIGURE 6

Hence $$T_{n,m} = \frac{n^{n+\frac{1}{2}}e^{-m}(\mu+m\zeta_0/\zeta)^m}{m!\,(n-m)^{n-m+\frac{1}{2}}n^m}\,e^{-(\mu+m\zeta_0/\zeta)}[1+O(n^{-1})]$$

and since $$\left(1-\frac{m}{n}\right)^{n-m+\frac{1}{2}} = e^{-m}[1+O(n^{-1})],$$

the appropriate limiting form of $T_{n,m}$ is

$$T_{n,m} = \frac{(\mu+m\zeta_0/\zeta)^m}{m!}\,e^{-(\mu+m\zeta_0/\zeta)}[1+O(n^{-1})]. \tag{20·2}$$

The majority of distributions encountered in practice are such that $\zeta_0 = 0$ for this condition implies that the chance that a thread exceeds S in strength tends to zero more rapidly than $1/S$ as S increases. In that case (20·2) degenerates to the Poisson formula

$$T_{n,m} = \frac{\mu^m}{m!}\,e^{-\mu}[1+O(n^{-1})]. \tag{20·3}$$

It is a useful confirmation of a later result to prove that when $\zeta > \zeta_0$ and $\zeta-\zeta_0$ is not small

$$\sum_{m=0}^{\infty}\frac{(\mu+m\zeta_0/\zeta)^m}{m!}\,e^{-(\mu+m\zeta_0/\zeta)} = \frac{1}{1-\zeta_0/\zeta}. \tag{20·4}$$

which is independent of μ.

The series converges since the ratio of the $(m+1)$th to the mth term tends to $\zeta_0/\zeta\,.\exp 1-\zeta_0/\zeta$ which is less than 1 except when $\zeta = \zeta_0$. It may be written in the form

$$-\sum_{m=0}^{\infty}\frac{1}{2\pi i}\int_C\frac{e^{-(\mu+m\zeta_0/\zeta)u}\,du}{(1-u)^{m+1}}$$

where the contour C encloses $u = 1$ and is such that $|\exp - u\zeta_0/\zeta| < |1-u|$ at every point on it. The curve $|1-u| = \nu\,|\exp - u\zeta_0/\zeta|$ satisfies this condition and has as one of its branches a closed loop containing $u = 1$, and also $u = 0$, provided that

$$1 < \nu < \zeta/\zeta_0\,.\exp(\zeta_0/\zeta - 1)$$

so C is chosen to be this loop. The series therefore has the sum

$$\frac{1}{2\pi i}\int_C\frac{e^{-\mu u}\,du}{u-1+e^{-u\zeta_0/\zeta}}.$$

The integrand has a simple pole inside the contour at which the residue is $1/(1-\zeta_0/\zeta)$ and some consideration will show that there are no other singularities within C. Consequently the sum of the series is $1/(1-\zeta_0/\zeta)$ when $\zeta > \zeta_0$, which is the required result.

Discrete limiting form of $T_{n,m,r}$

21. The corresponding discrete limiting form of

$$T_{n,m,r} = \frac{(n-r)!}{(m-r)!\,(n-m)!}\,(\kappa b_m)^{n-m}(1-\kappa b_m)^{m-r} \quad (0 \leqslant \kappa < 1/b_r),$$

is appropriate when n, m and r are all large, but $m-r$ is $O(1)$. Retaining the notation of the previous sections, we expand $a(z/\zeta)$ near x as

$$a(z/\zeta) = a(x/\zeta) + \frac{(z-x)}{\zeta} a'(x/\zeta) + O(z-x)^2$$

i.e.
$$b_m = b_r\left[1 - \frac{(m-r)}{n}\frac{\zeta_x}{\zeta}\right] + O(n^{-2}) \tag{21·1}$$

where $\zeta_x = -\dfrac{a'(x/\zeta)}{a(x/\zeta)}$, and $1/\zeta_x$ is the subtangent to the $a(w)$ curve at $w = x/\zeta$. The important values of κ are near its maximum $1/b_r$ and accordingly we write

$$\kappa = \frac{1}{b_r}\left(1 - \frac{\mu}{n}\right)$$

or in terms of $\lambda = \kappa(1-x)$,
$$\lambda = \lambda_x\left(1 - \frac{\mu}{n}\right). \tag{21·2}$$

Then
$$\begin{aligned}(n-m)\log \kappa b_m &= -\frac{(n-m)}{n}[\mu + (m-r)\zeta_x/\zeta] + O(n^{-1})\\ &= -(1-x)[\mu + (m-r)\zeta_x/\zeta] + O(n^{-1}),\\ 1 - \kappa b_m &= \frac{1}{n}[\mu + (m-r)\zeta_x/\zeta] + O(n^{-2})\end{aligned}$$

and $T_{n,m,r}$ reduces to

$$T_{n,m,r} = \frac{(1-x)^{m-r}[\mu + (m-r)\zeta_x/\zeta]^{m-r}}{(m-r)!} e^{-(1-x)[\mu+(m-r)\zeta_x/\zeta]}[1 + O(n^{-1})]. \tag{21·3}$$

It follows immediately from (20·4) that, to $O(n^{-1})$,

$$\sum_{m=r}^{\infty} T_{n,m,r} = \frac{1}{1 - \dfrac{(1-x)\zeta_x}{\zeta}} = \frac{1}{1 + \dfrac{(1-x)}{\zeta}\dfrac{a'(x/\zeta)}{a(x/\zeta)}}, \tag{21·4}$$

provided that the denominator is not small, or in other words, that x is not near a critical point.

Limiting form of $Q_{n,m}$ for large n

22. One can now examine the asymptotic behaviour of $Q_{n,m}$ and ultimately that of $B_n = Q_{n,0}$ by means of the identity (13·5), viz.

$$\sum_{m=r}^{n} Q_{n,m} \frac{(n-r)!\,(\kappa b_m)^{n-m}(1-\kappa b_m)^{m-r}}{(m-r)!\,(n-m)!} \equiv 1 \tag{13·5}$$

$$0 \leqslant r \leqslant n, \quad 0 \leqslant \kappa \leqslant \frac{1}{b_n}.$$

The limiting form of (13·5) depends on the situation of the roots of

$$1-z-\lambda a(z/\zeta)=0$$

which lie in the range $x\leqslant z\leqslant 1$, where, as defined above,

$$\lambda=\kappa(1-x),\quad 0\leqslant\lambda\leqslant\lambda_x=(1-x)/a(x/\zeta).$$

The discussion is first confined to what we call the 'non-critical' regions where z is not near a critical point. Suppose, for simplicity, that the $a(w)$ curve is of the type shown in figure 7 having at most 3 positive real roots. As before, λ_c denotes the critical value of λ at which the two largest roots z_2, z_3 coincide at z_c for fixed ζ, the smallest root then being $z_{c'}$.

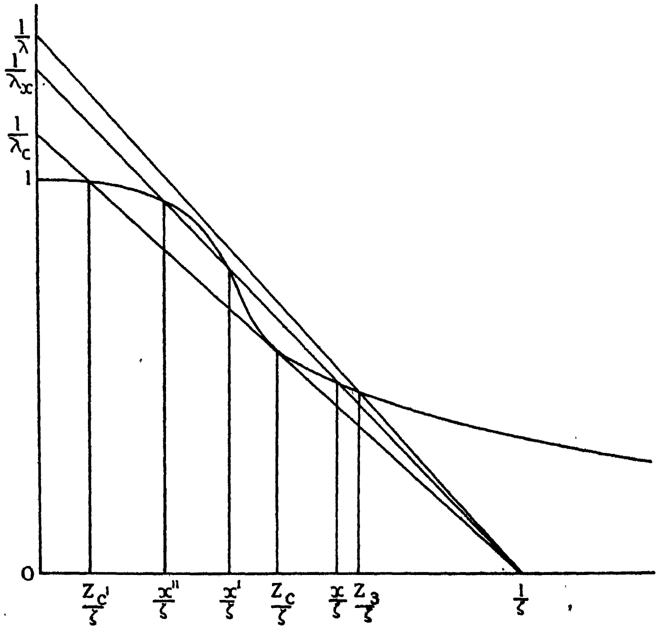


FIGURE 7

Consider what happens as x changes from 1 to 0, ζ being fixed, confining attention to the non-critical regions.

(i) $z_c<x<1$. When x is in this range there is only one root z_3 in $x\leqslant z\leqslant 1$ and if neither z_3-x nor $1-z_3$ are small, (16·1) may be substituted in (13·5) to give

$$\int_{-\infty}^{\infty} Q_n(z)\,dz\,\sqrt{\left(\frac{n(1-x)}{2\pi(z_3-x)(1-z_3)}\right)}$$
$$\times\exp\left[-\frac{n(1-x)\left[1+\frac{\lambda}{\zeta}a'(z_3/\zeta)\right]^2(z-z_3)^2}{2(z_3-x)(1-z_3)}\right][1+A(z-z_3)+Bn(z-z_3)^3]$$
$$=1+O(n^{-1}) \qquad (22·1)$$

where $Q_n(z)$ is written for $Q_{n,m}$.

On the other hand, when $z_3 - x$ is small so that $m-r = O(1)$ and $\lambda = \lambda_x\left(1-\frac{\mu}{n}\right)$ with $\mu = O(1)$ the appropriate limiting form is, from (21·3),

$$\sum_{m=r}^{\infty} Q_{n,m} \frac{[(1-x)(\mu+(m-r)\zeta_x/\zeta)]^{m-n}}{(m-r)!} e^{-(1-x)(\mu+(m-r)\zeta_x/\zeta)} = 1+O(n^{-1}) \qquad (22\cdot2)$$

where $\zeta_x = -a'(x/\zeta)/a(x/\zeta)$.

It is now proved that when $z_c < z < 1$ and $z - z_c = O(1)$,

$$Q_n(z) = \left\{1 + \frac{(1-z)\,a'(z/\zeta)}{\zeta a(z/\zeta)}\right\}(1+O(n^{-1})). \qquad (22\cdot3)$$

The values $\mu = 0$ and $\mu = \zeta_x/\zeta$ in (22·2) give respectively

$$\sum_{m=r}^{\infty} Q_{n,m} \frac{[(1-x)(m-r)\zeta_x/\zeta]^{m-r}}{(m-r)!} e^{-(1-x)(m-r)\zeta_x/\zeta} = 1+O(n^{-1}),$$

$$\sum_{m=r}^{\infty} Q_{n,m} \frac{[(1-x)(m-r+1)\zeta_x/\zeta]^{m-r}}{(m-r)!} e^{-(1-x)(m-r+1)\zeta_x/\zeta} = 1+O(n^{-1}).$$

Writing $r+1$ for r and multiplying through by $(1-x)\,\zeta_x/\zeta$ the latter becomes

$$\sum_{m=r+1}^{\infty} Q_{n,m} \frac{[(1-x)(m-r)\zeta_x/\zeta]^{m-r}}{(m-r)!} e^{-(1-x)(m-r)\zeta_x/\zeta} = (1-x)\,\zeta_x/\zeta(1+O(n^{-1}))$$

whence, on subtraction,

$$Q_{n,r} = Q_n(x) = \{1-(1-x)\,\zeta_x/\zeta\}(1+O(n^{-1}))$$

which is equivalent to (22·3). The proof can be made rigorous by inserting the approximation factor $\exp A(m-r)^2/n$ in place of $1+O(n^{-1})$ and confining the summation to the range in which this factor does not begin to dominate, the remainder being demonstrably negligible.

The same expression for $Q_n(z)$ is seen to satisfy (22·1), as it should, for then, to $O(n^{-1})$,

$$Q_n(z_3) = 1 + \frac{(1-z_3)}{\zeta}\frac{a'(z_3/\zeta)}{a(z_3/\zeta)} = 1 + \frac{\lambda}{\zeta}a'(z_3/\zeta)$$

and expanding $Q_n(z)$ in (22·1) as $Q_n(z_3) + Q_n'(z_3)(z-z_3) + O(z-z_3)^2$, the first term integrates to 1, the odd powers of $z - z_3$ vanish on integration, and $O(z-z_3)^2 = O(n^{-1})$ in the effective range of $z - z_3$.

It also follows that over the important range of m in (22·2), $Q_{n,m}$ differs from $Q_n(x)$ only by $O(n^{-1})$ and on substituting this value for $Q_{n,m}$ the result (21·4) is verified.

(ii) $z_{c'} < x < z_c$. Avoiding for the present the critical region near z_c, we next discuss the range between $z_{c'}$ and z_c. There are two cases typified by the values x' and x'' in figure 7, both corresponding to $\lambda = \lambda_x$ as shown.

Consider first the case of x'. When $\lambda < \lambda_x$ the only root greater than x' is z_3, which is also greater than x, and (22·1) holds with x' for x. On making $\lambda = \lambda_x$, however, the roots are x' and x, and (13·5) becomes

$$Q_{n,r'}T_{n,r',r'} + Q_{n,r'+1}T_{n,r'+1,r'} + \ldots + \int_{-\infty}^{\infty} Q_n(z)\,dz \sqrt{\left(\frac{n(1-x')}{2\pi(x-x')(1-x)}\right)}$$
$$\times \exp\left[-\frac{n(1-x')\left[1+\frac{\lambda_x}{\zeta}a'(x/\zeta)\right]^2(z-x)^2}{2(x-x')(1-x)}\right][1+A(z-x)+Bn(z-x)^3]$$
$$= 1+O(n^{-1}), \tag{22·4}$$

where $\frac{r'}{n} = x'$, and it is sufficient to note that $T_{n,r',r'} = 1+O(n^{-1})$.

The integral reduces to $1+O(n^{-1})$ by virtue of (22·3), and so

$$Q_{n,r'} = Q_n(x') = O(n^{-1}). \tag{22·5}$$

In the case of x'', when we make $\lambda = \lambda_x$ there are three roots at x'', x' and x; then (13·5) takes a form similar to (22·4) but with r'' for r', and two integrals in the neighbourhoods of x' and x respectively. The first integral vanishes by (22·5), the second is $1+O(n^{-1})$, and as before $Q_{n,r''} = Q_n(x'') = O(n^{-1})$.

When $x''-x'$ is small the discussion is a little more elaborate but the same result is true. Hence for any value of z in the non-critical region between $z_{c'}$ and z_c,

$$Q_n(z) = O(n^{-1}). \tag{22·6}$$

(iii) $0 \leqslant x < z_{c'}$. As x'' passes through the value $z_{c'}$ into the range $0 \leqslant x'' < z_{c'}$, the other roots x', x for $\lambda' = \lambda_x$ coalesce and become complex, so by the previous argument $Q_n(z)$ again satisfies (22·3).

The conclusions are therefore summarized as follows. Let $a(w)$ be such that $1-z-\lambda a(z/\zeta) = 0$ has at most 3 positive real roots, and let ζ be chosen so that when z_2 and z_3 are made to coalesce at z_c for suitable $\lambda = \lambda_c$ the smallest root z_1 has a value $z_{c'} > 0$. Then the approximate form of $Q_n(z)$ in the non-critical regions of z is

$$\left.\begin{aligned} Q_n(z) &= \left\{1+\frac{(1-z)\,a'(z/\zeta)}{\zeta a(z/\zeta)}\right\}(1+O(n^{-1})) && \text{in } z_c<z<1,\\ &= O(n^{-1}) && \text{in } z_{c'}<z<z_c,\\ &= \left\{1+\frac{(1-z)\,a'(z/\zeta)}{\zeta a(z/\zeta)}\right\}(1+O(n^{-1})) && \text{in } 0\leqslant z<z_{c'}.\end{aligned}\right\} \tag{22·7}$$

By varying ζ the range $z_{c'}<z<z_c$ may be made to vanish or to include $z = 0$ but in all cases (22·6) holds throughout any part of this range, and (22·3) otherwise (excluding values of z near z_c or $z_{c'}$). The argument can evidently be extended to forms of $a(w)$ where the maximum number of roots is greater than 3.

23. The result just obtained has an immediate application to $B_n = Q_{n,0}$. Let $\lambda = 1$ so that the lowest root $z_1 = 0$ and allow ζ to vary, the critical value ζ_r corresponding to the double root z_r. Suppose $a(w)$ to have the form shown in figure 8.

The remaining non-zero roots z_2, z_3 are real or complex according as $\zeta<\zeta_\tau$ or $\zeta>\zeta_\tau$. Putting $z=0$, $a(0)=1$, $a'(0)=-\zeta_0$ in (22·7), it can therefore be stated that

$$\left.\begin{aligned} B_n &= Q_{n,0} = O(n^{-1}) && \text{when } \zeta<\zeta_\tau, \\ B_n &= (1-\zeta_0/\zeta)(1+O(n^{-1})) && \text{when } \zeta>\zeta_\tau, \end{aligned}\right\} \tag{23·1}$$

provided $$|\zeta-\zeta_\tau| = O(1).$$

Figure 9 illustrates the special case when $\zeta_\tau=\zeta_0$ and there is only one non-zero root.

The fact that $B_n \sim 1-\zeta_0/\zeta$ when $\zeta>\zeta_\tau$ is perhaps not unexpected in view of the similar exact expression for B_n (12·2) which holds for a form of $a(w)$ to which the present more general case approximates near $w=0$. But the result (23·1) sheds an interesting light on the argument and conclusions reached about large bundles in the early part of the paper (§§ 4, 5, 6). It was deduced there that all very large bundles of a given size n would break at loads approximating to $n\zeta_\tau$ in the present notation, provided that a unique value of ζ_τ exists. Evidently this is not the whole truth of the matter. If the chance $1-b(s)$ of a thread exceeding s in strength tends to zero at a rate proportional to $1/s$, then, however large the bundle may be, its strength $S=n\zeta$ can take values ranging from $n\zeta_\tau$ to ∞ with probabilities according to (23·1). In such cases (23·1) provides all the information necessary about B_n for large n, since the small excluded region near ζ_τ is negligible compared with the wide range over which B_n varies slowly with ζ. In practice, however, it is generally found that $1-b(s)$ tends to zero more rapidly than $1/s$ so that $\zeta_0=0$, and (23·1) then gives no information about B_n in the region where it differs appreciably from 0 or 1.

24. Finally the behaviour of $Q_{n,m}$ in the critical region is discussed. Keeping $\lambda=1$, $z_1=0$, let ζ now approach ζ_τ. The two largest roots z_2, z_3 come into the vicinity of z_τ, and on substituting (18·7) and (20·2) the appropriate limiting form of (13·5) is found to be

$$\sum_{m=0}^{\infty} Q_{n,m}\frac{(m\zeta_0/\zeta)^m}{m!}e^{-m\zeta_0/\zeta}+\int_{-\infty}^{\infty} Q_n(z)\,dz\,\sqrt{\left(\frac{n}{2\pi z_\tau(1-z_\tau)}\right)} \times\exp\left\{-\frac{n}{2z_\tau(1-z_\tau)}\left[\frac{a''}{2\zeta_\tau^2}(z-z_\tau)^2+\frac{z_\tau}{\zeta_\tau}(\zeta-\zeta_\tau)\right]^2\right\} = 1+O(n^{-\frac{1}{2}}) \tag{24·1}$$

where $$\zeta-\zeta_\tau = O(n^{-\frac{1}{2}}).$$

The most commonly occurring type of thread population has $\zeta_0=0$ and (24·1) then reduces to

$$B_n = Q_{n,0} = 1-\int_{-\infty}^{\infty} Q_n(z)\,dz\,\sqrt{\left(\frac{n}{2\pi z_\tau(1-z_\tau)}\right)} \times\exp\left\{-\frac{n}{2z_\tau(1-z_\tau)}\left[\frac{a''}{2\zeta_\tau^2}(z-z_\tau)^2+\frac{z_\tau}{\zeta_\tau}(\zeta-\zeta_\tau)\right]^2\right\} = 1+O(n^{-\frac{1}{2}}). \tag{24·2}$$

When $\zeta_0\neq 0$ it can be shown by an argument similar to that used to prove (22·3) that the effect is to multiply the right-hand side of this equation by a factor $1-\zeta_0/\zeta$.

The behaviour of B_n for values of ζ satisfying $\zeta - \zeta_T = O(n^{-\frac{1}{2}})$ is thus seen to depend on that of $Q_n(z)$ over a range of z such that $z - z_T = O(n^{-\frac{1}{4}})$. Therefore consider

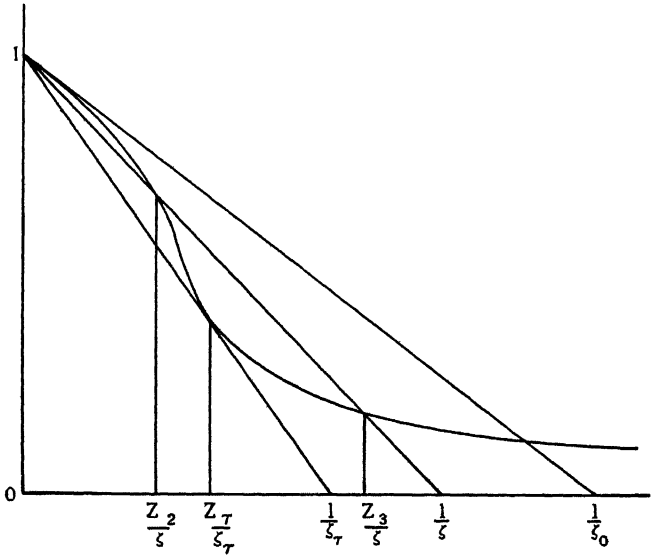


FIGURE 8

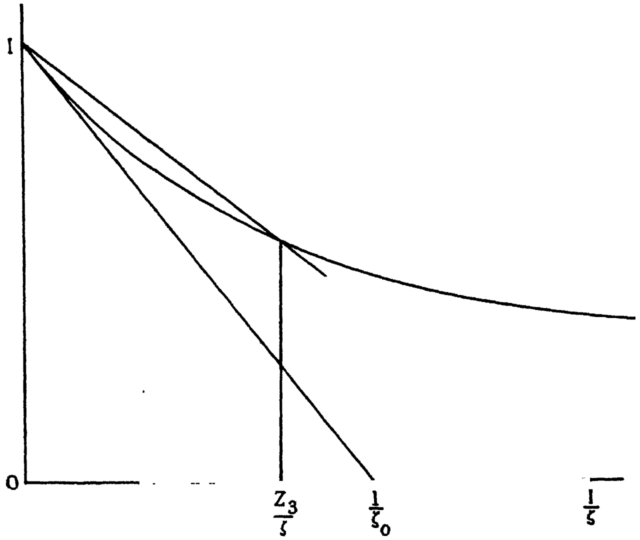


FIGURE 9

the limiting form of (13·5) when x is in the critical region and $z_3 - z_2 = O(n^{-\frac{1}{4}})$. It is more convenient for the moment to work in terms of z_c. Suppose first that $z_c < x$, and now make use of the result (19·2) that near z_3, $T_n(z, x)$ take the form appropriate to an isolated simple root, but to the order $O(n^{-\frac{1}{4}})$ only. Since z_3 is the only root in $x<z<1$, (13·5) becomes

$$\int_{-\infty}^{\infty} Q_n(z)\,dz \bigg/ \sqrt{\left(\frac{n}{2\pi(z_3-x)}\right)} \exp\left[-\frac{n\lambda_c^2 a_c''^2}{2\zeta^4}\frac{(z_3-z_c)^2}{(z_3-x)}(z-z_3)^2\right]$$
$$\times[1+An^{\frac{1}{4}}(z-z_3)+Bn(z-z_3)^3] = 1+O(n^{-\frac{1}{4}}) \qquad (24\cdot3)$$

and this is satisfied by $$Q_n(z) = \frac{\lambda_c a_c''}{\zeta^2}(z-z_c)[1+O(n^{-\frac{1}{4}})] \qquad (24\cdot4)$$

which is the same as (22·3) to $O(n^{-\frac{1}{4}})$. Such an argument does not establish (24·4) as the unique solution of (24·3), but it is possible to construct a direct rigorous proof on the lines of § 22 (i), though the details are tedious.

By taking $x<z_c$, $z_c-x = O(n^{-\frac{1}{4}})$, the behaviour of $Q_n(z)$ when $z<z_c$, $z_c-z = O(n^{-\frac{1}{4}})$ may be studied by a similar extension of § 22 (ii), and it is found that

$$Q_n(z) = O(n^{-\frac{1}{4}}) \qquad (24\cdot5)$$

where $$z<z_c, \quad z_c-z = O(n^{-\frac{1}{4}}).$$

The asymptotic form of B_n

25. Applying these formulae to (24·2) one writes to $O(n^{-\frac{1}{4}})$

$$\lambda_c = 1, \quad z_c = z_\tau, \quad \lambda_c a_c''/\zeta^2 = a_\tau''/\zeta_\tau^2.$$

The contribution to the integral in (24·2) from points in a range of $z-z_\tau$ smaller than $O(n^{-\frac{1}{4}})$ can be shown to be negligible to $O(n^{-\frac{1}{4}})$ and it is concluded that

$$B_n = 1-\int_0^\infty \frac{a_\tau''}{\zeta_\tau^2}(z-z_\tau)\,dz \bigg/ \sqrt{\left(\frac{n}{2\pi z_\tau(1-z_\tau)}\right)}$$
$$\times\exp\left\{-\frac{n}{2z_\tau(1-z_\tau)}\left[\frac{a_\tau''}{2\zeta_\tau^2}(z-z_\tau)^2+\frac{z_\tau}{\zeta_\tau}(\zeta-\zeta_\tau)\right]^2+O(n^{-\frac{1}{4}})\right\}$$

which reduces on setting

$$\frac{z_\tau}{\zeta_\tau}y = \frac{a_\tau''}{2\zeta_\tau^2}(z-z_\tau)^2+\frac{z_\tau}{\zeta_\tau}(\zeta-\zeta_\tau)$$

to $$B_n = \int_{-\infty}^{\zeta-\zeta_\tau}\frac{dy}{\zeta_\tau}\sqrt{\left(\frac{nz_\tau}{2\pi(1-z_\tau)}\right)}\exp\left[-\frac{nz_\tau y^2}{2\zeta_\tau^2(1-z_\tau)}\right]+O(n^{-\frac{1}{4}}). \qquad (25\cdot1)$$

That is, the probability distribution of ζ tends to the normal form for large n with expectation ζ_τ and standard deviation

$$\sigma = \zeta_\tau\sqrt{\left(\frac{1-z_\tau}{nz_\tau}\right)} = \zeta_\tau\sqrt{\left(\frac{a_\tau}{n(1-a_\tau)}\right)}.$$

When ζ_0 is not zero the formula for B_n is simply multiplied by $1-\zeta/\zeta_0$.

The final result restated in terms of the earlier notation is therefore as follows. If all threads have the same load-extension curve, and the chance $b(s)$ of a thread breaking under load s is such that $1-b(s)$ tends to zero faster than $1/s$, then the strength S of a bundle of n threads has a probability distribution which tends for large n to the normal form with expectation

$$S_\tau = ns_\tau[1-b(s_\tau)]$$

and standard deviation $\qquad \sigma = s_\tau\sqrt{\{nb(s_\tau)\,[1-b(s_\tau)]\}},$

where s_τ gives $s[1-b(s)]$ its greatest value.

It is worth remarking that the same result would follow at once if one could assume that for sufficiently large n the probability distribution of breaking extension is independent of that of the number of threads surviving up to the point of rupture, the probability of this number being then distributed according to a simple binomial law. Subsequent work on breaking extension suggests that the assumption is in fact true, but there seems to be no *a priori* justification for it.

Note also that I have assumed the dispersion of α to be negligible compared to that of ζ, but when n is large it may become of comparable order, in which case the formulae are not strictly valid.

Part II of this paper will contain a generalization to threads having varying load-extension curves, a discussion of the probability distribution of extension at break, and a further investigation into the asymptotic form of B_n by a more powerful saddle-point method based on the special properties of B_n when $b(s) = e^{-\kappa/s}$.

My thanks are due to the Director and Council of the Wool Industries Research Association for permission to publish this work.

REFERENCES

Peirce, F. T. 1926 *J. Textile Inst.* **17**, T. 355. (See also preceding papers of the series, with O. Midgley.)

Weibull, W. 1939 *IngenVetenskAkad. Handl.* no. 153 (in English.)

Stationary electron swarms in electromagnetic fields

By D. Gabor, Dr.Ing., F.Inst.P.

(*Communicated by S. R. Milner, F.R.S.—Received* 12 *June* 1944)

Electron clouds rotating in axially symmetric magnetic fields have been known for a long time, but the agreement between theory and experiment is still very unsatisfactory. The discrepancy appears to be due to the interaction of electrons. Before approaching this difficult problem it is desirable to possess a more complete theory of stationary swarms without interaction. In the present paper the distribution density is calculated on the basis of classical statistical mechanics. It is shown that electrons injected at any point with very small initial velocities will distribute themselves with a density inversely proportional to the distance from the axis, in a certain annular space. Only the limits of this space, not the distribution inside it, will be dependent on the electric or magnetic fields. The uniform or nearly uniform distributions calculated by previous authors are singular solutions, inconsistent with any degree of statistical disorder. Other laws of density distribution can be realized by simultaneous injection of electrons at several points. These offer a possibility to realize dispersing electron lenses and corrected electron optical systems. It is shown that the ring current produced by the rotating electron cloud can reduce the magnetic field at the axis very considerably in devices of practicable dimensions. It appears also possible to produce clouds of free electrons with densities sufficient for observable optical effects.

The theory of the stationary motion of electron assemblies in electric and magnetic fields is of interest in two fields of applied electronics. It arises in the theory of the magnetron, an electronic device which so far has defied all attempts at a quantitative theoretical explanation. The second field is electron optics, which suffers from the fundamental limitation that lens correction is impossible so long as only electromagnetic fields free of space charge are employed. Hence the application of electron clouds offers one of the very few prospects for further development of electron optical devices, especially of the electron microscope.

The full-anode magnetron, a diode with a cylindrical anode and a coaxial filament as cathode, was first constructed and investigated by A. W. Hull in 1920, and has been explored theoretically and experimentally by numerous authors.* An especially thorough theoretical study is due to L. Brillouin (1941, 1942). These investigations have revealed a striking discrepancy between theory and the experimental results.† Hull's simple theory led to the conclusion that current could flow to the anode only above a certain critical voltage, which increases quadratically with the magnetic field intensity. Below this voltage the current ought to be cut off, and an electron cloud of uniform charge density

$$-\rho_H = \frac{eH^2}{8\pi mc^2} \tag{1}$$

should rotate in the magnetron like a solid body, with an angular velocity

$$\omega_H = \frac{eH}{2mc}, \tag{2}$$

---

* A. W. Hull (1921, 1924). F. B. Pidduck (1936). Extensive bibliography in A. F. Harvey (1943).

† See especially Harvey (1943, pp. 83–113) and also the Foreword by E. B. Moullin.

which is one-half of the 'Larmor frequency'. Though the experiments confirmed the existence of a critical voltage, the quantitative agreement was very unsatisfactory. Current starts flowing considerably below the critical voltage and reaches its saturation value only very gradually, whereas on Hull's theory, even making an allowance for the initial velocities of the electrons, the rise ought to be confined within a few tenths of a volt. Though later workers added some refinements to Hull's theory, they were unable to account for the width and shape of the cut-off characteristic. An exception is an investigation by E. G. Linder (1938 *a*, *b*), who found experimentally that in a rotating electron cloud random motion develops with a very high electron temperature, of the order of 10 V. On this basis he could account for the width of the cut-off curves.

Several objections could be raised against Linder's results. He finds by probe measurements a particularly high electron temperature when the current is entirely cut off, which is a thermodynamical impossibility. His method of calculating the current from an electron cloud to a plate is also open to objection, as he uses a formula which is strictly valid only in the absence of a magnetic field. Nevertheless, it appears very likely that Linder has correctly located the root of the discrepancy. The new difficulty arises, however, that theory cannot account for the experimentally found rapid development of random motion in an electron cloud. This phenomenon is closely related to the inexplicably rapid establishment of electron temperature in gas discharges at very low pressures, which was discovered by I. Langmuir and H. Mott-Smith, and discussed in detail by Langmuir (1927, 1928). As the problem of electron interaction is one of extraordinary mathematical difficulty, it appears desirable to obtain first a reliable experimental estimate of the effect. The first step towards this must be a more complete investigation of the problem *without* interaction. It will be shown that no satisfactory basis for this has existed up to now, as the corrections which have to be applied to Hull's results are very much larger than those found by any of the later authors who tried to improve on them.

Hull's results (1) and (2) can be derived from the condition of zero radial current. In general a zero current would be the resultant of two equal and opposite currents, one flowing inwards, the other outwards. If the electrons start from the cathode with zero initial velocity, as assumed in the simple theory, reversal of the sense of motion can take place only at the outer boundary of the electron cloud. By formulating the condition of equal and opposite currents, combining this with the dynamical equations and substituting it into Poisson's equation, a differential equation for the space-charge distribution is obtained.* The remarkable result follows, that this equation has a solution free of singularities only in the case when both the opposite currents are zero, that is to say, if there is no radial electron motion at all. This means that one must imagine the electrons circling around the cathode in coaxial circular orbits. The question immediately arises how such a state of motion could ever establish itself. Some authors have thought that it could be justified, as they found that small radial velocities produce only insignificant deviations from Hull's law of distribution. It will be shown, however, that any initial velocities, however slight, require a fundamental modification of the picture, and that Hull's results must be considered as a singular solution of the problem, which could neither establish, nor maintain itself.

---

* Cf. e.g. Brillouin (1941).

This is of great importance from the point of view of the other field mentioned at the beginning, the electron optical applications of space charges. Electron optical systems cannot be corrected at present for spherical aberration and other defects, because no dispersing electron lenses can be realized without (negative) space charges. A space charge according to equation (1) would be useless for this purpose. As explained above, this distribution satisfies the condition that electrons are everywhere in radial equilibrium. Hence, if an electron beam is shot through such an electron cloud, the concentrating effect of the magnetic field would exactly balance the dispersing effect of the space charge, and the resulting lens effect would be *nil*. But, as it will be shown that stationary electron clouds are possible in which radial equilibrium exists only in the average, but not at every radius, the prospect of electron optical applications is not closed. The problem will therefore be formulated in sufficient generality to cover both the magnetron and the electron optical applications.

The general method to be followed will be to treat the problem as one of statistical mechanics. First the distribution law will be derived on the basis of classical statistics, giving the density as a function of the position and of the electromagnetic potentials. The second step will be to substitute the density into Poisson's equation and find the density distribution, and the potential consistent with it, as a function of position. Fortunately, the first step leads to an extremely simple law, so that the second step and the discussion of the solution will present no difficulty.

1. THE DYNAMICAL PROBLEM

Assume an axially symmetrical electromagnetic field, specified in cylindrical co-ordinates (z, r, θ) by an electrostatic potential $\phi(z, r)$ and a vector potential $A(z, r)$. As the vector potential has only a tangential component, it will not be necessary to write it as a vector, or to distinguish it by a suffix t like the other tangential components. The units will be Gaussian. The magnetic field follows from A by the relations

$$H_r = \operatorname{curl}_r A = -\frac{\partial A}{\partial z}, \quad H_z = \operatorname{curl}_z A = \frac{1}{r}\frac{\partial}{\partial r}(rA). \tag{3}$$

In the special case of a homogeneous magnetic field, $H_r = 0$, $H_z = H_0$

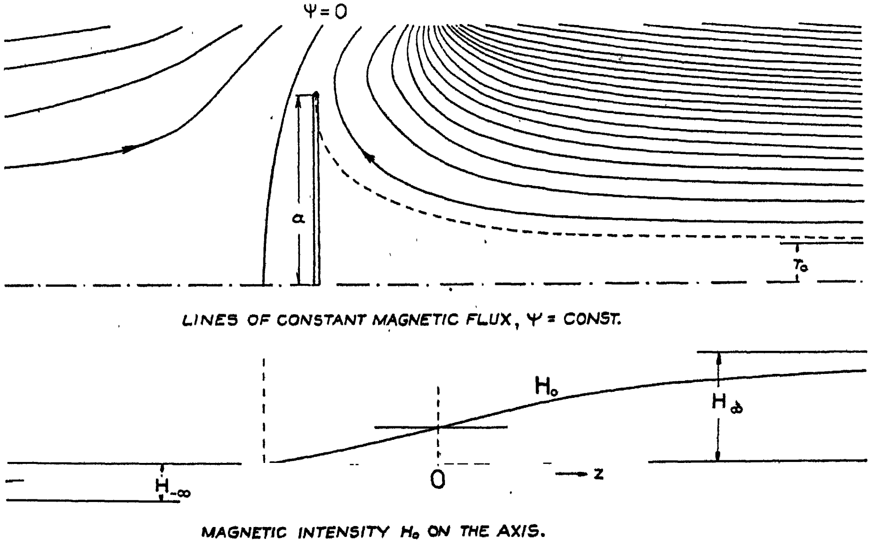
$$A = \tfrac{1}{2}H_0 r + C/r, \tag{4}$$

where C is an arbitrary constant. As it does not figure in the dynamical equations it can be put zero from the start.

Now assume a cathode in the form of a circular filament with negligible thickness, with a radius a at a position where the vector potential has the value A_0. Figure 1 illustrates an example. The cathode is placed in, or near to, a position of zero magnetic flux. In the case illustrated, the field at the axis has the law

$$H(0,z) = \text{const.}\,(\tfrac{1}{2}+\tan^{-1}z),$$

so that at great distance from the cathode, in either direction, the field becomes homogeneous, the field at the right being three times stronger than the field at the left, which plays only the part of an auxiliary, with the purpose of adjusting A_0 to a suitable (small) value. The magnetic field is illustrated by the field lines, i.e. the meridians of the tubes of constant flux $\Psi = 2\pi rA = \text{const}$. The flux increases by equal steps from one line to the next. The field line passing through the cathode is shown in dotted lines. It approaches a radius r_0 at the far right, where the field becomes homogeneous. This arrangement is suitable for electron optical applications, as it leaves the axis free. The cathode has to be placed at or near the line $\Psi = 0$, as otherwise the electrons could not get near the axis.



LINES OF CONSTANT MAGNETIC FLUX, Ψ = CONST.

MAGNETIC INTENSITY H_0 ON THE AXIS.

FIGURE 1. Example of magnetic field.

The dynamical equations are most easily formulated in the Hamiltonian form. The linear momentum of the electron in an electromagnetic field is defined as

$$\mathbf{p} = m\mathbf{v}-\frac{e}{c}\mathbf{A}. \tag{5}$$

In an electromagnetic field of rotational symmetry it is advantageous to write the components of the linear momentum p_z, p_r and p_θ/r, where p_θ is the angular momentum. The Hamiltonian is

$$\mathscr{H} = \frac{1}{2m}\left[p_z^2+p_r^2+\left(\frac{1}{r}p_\theta+\frac{e}{c}A\right)^2\right]-e\phi. \tag{6}$$

The potential ϕ is to be measured from the cathode as zero level.

* R. Becker (1933, p. 96). The equation (5) is due to K. Schwarzschild (1903).

The problem has two integrals, $\mathscr{H} = \text{const.}$ and $p_\theta = \text{const.}$ Both can be immediately obtained from the canonical equations. Marking values relating to the cathode with a suffix 0, the momentum integral is written

$$p_\theta = r\left(mv_t - \frac{e}{c}A\right) = a\left(mv_{t0} - \frac{e}{c}A_0\right). \tag{7}$$

From this point on it will be convenient to measure the velocity in potential units. Therefore v is replaced by

$$u = (\sqrt{[m/2e]})\,v. \tag{8}$$

It will also be convenient to introduce a quantity, which could be called the vector potential relative to the cathode, defined by

$$\mathscr{A} = \sqrt{\left(\frac{e}{2mc^2}\right)}\left(A - \frac{a}{r}A_0\right). \tag{9}$$

$r\mathscr{A}$ is a measure of the flux which passes between the cathode (a, z_0) and the circle (r, z). $\mathscr{A}$ has the same dimensions as u. With these new symbols equation (7) becomes

$$ru_t - au_{t0} = r\mathscr{A}. \tag{7·1}$$

The energy integral is, in the new units,

$$u_z^2 + u_r^2 + u_t^2 - \phi = (u_r^2 + u_z^2 + u_t^2)_0. \tag{10}$$

A condition which electrons have to fulfil in order to have access to an element in phase space can now be formulated by eliminating u_{t0} between the equations (7·1) and (10). It will be convenient to introduce a dimensionless parameter

$$x = r/a \tag{11}$$

for the radius, i.e. to measure r in units a. Thus

$$Q \equiv u_z^2 + u_r^2 - (x^2-1)\left[u_t - \frac{x^2}{x^2-1}\mathscr{A}\right]^2 - \left[\phi - \frac{x^2}{x^2-1}\mathscr{A}^2\right] = u_{z0}^2 + u_{r0}^2 > 0. \tag{12}$$

This is the *criterion of accessibility*. Its geometrical interpretation in velocity space is shown in figure 2. $Q = 0$ is a quadric surface of rotational symmetry relative to the u_t-axis. Its type is given by the following table:

x	<1	>1
$\phi > x^2\mathscr{A}^2/(x^2-1)$	real ellipsoid	one-shell hyperboloid
$\phi < x^2\mathscr{A}^2/(x^2-1)$	imaginary ellipsoid	two-shell hyperboloid

This classification is illustrated in figure 3 in the special case of a homogeneous magnetic field, in which

$$\mathscr{A} = K\left(x - \frac{1}{x}\right), \quad \frac{x^2}{x^2-1}\mathscr{A}^2 = K^2(x^2-1), \tag{13}$$

where $K = a\sqrt{(eH^2/8mc^2)}$.

Now consider electrons which have left the cathode in the velocity interval u_0, u_0+du_0. They will arrive at some point (z, r) with a velocity u, and in an interval du, connected with the initial data by the relations

$$u^2 = u_0^2+\phi, \quad u\,du = u_0\,du_0. \tag{14}$$

Combining this with equation (12) it is seen that electrons which have access to (z, r) will all be contained in that segment of the spherical shell with radii u, $u+du$ which is (seen from the axis) *outside* the quadric surface $Q = 0$. This shell segment is the accessible volume in velocity space. It may be noted that it gives also a measure of the volume of momentum space, as, by equation (5),

$$\frac{\partial(p_x, p_y, p_z)}{\partial(v_x, v_y, v_z)} = m^3 = \text{const.} \tag{15}$$

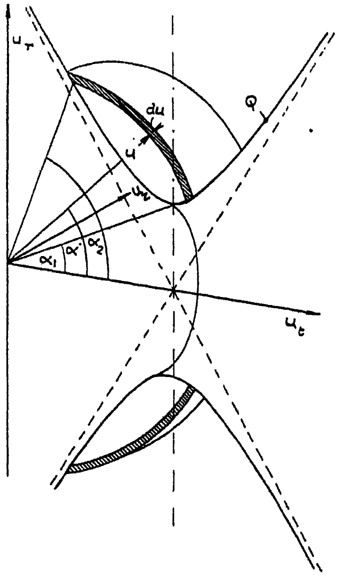


FIGURE 2. Construction of the accessible volume of momentum space.

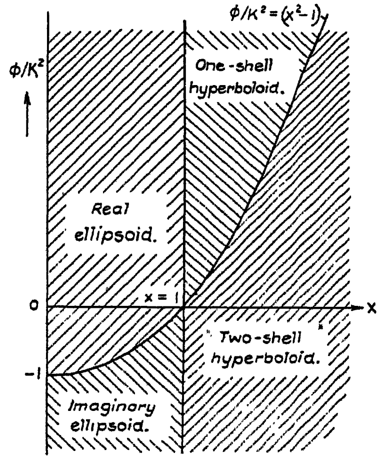


FIGURE 3. The quadric surface Q.

2. THE LAW OF VELOCITY DISTRIBUTION

By Liouville's theorem the density D of a group of electrons in phase space is the same as the original density which this group had at the cathode. Hence if it were known that the whole accessible volume of momentum or (velocity) space calculated in the previous section would be actually filled with electrons, then the velocity distribution at any point would also be known, and from this the density in configuration space could be calculated. This condition is fulfilled in an important special case, in that of the infinitely long magnetron with cylindrical cathode, the only one considered by previous authors. In this case there is no further limitation of the accessible volume, as it is easy to see that a trajectory continued backwards from any point of it must actually reach the cylindrical cathode at some point.

In the general case Liouville's theorem by itself is not sufficient to determine the velocity distribution, for two reasons. First, in the case of limited cathodes in general, only a part of the accessible volume will be really filled. This difficulty arises especially in the case of an 'ordered' flow of electrons. Hull's distribution with electrons circling in circular orbits is an extreme example of such an 'ordered' flow. But this uncertainty will exist also in all cases of motion in 'short trajectories', i.e. if the electrons return to the cathode after spending only a short time in the cloud. There is, however, also a second sort of difficulty, which arises especially in the case of 'long trajectories'. The same group of electrons might return—and by the theorem of Poincaré and Zermelo *will* return—an indefinite number of times to the same element in phase space, if it is allowed to spend a long time in the cloud, and the number of times this will happen is not known. Both difficulties are resolved if this time can be assumed to be very long, by the *ergodic theorem*—long known as the ergodic hypothesis, until G. D. Birkhoff and J. v. Neumann proved it—according to which the whole accessible phase volume will be filled, with uniform density.* Moreover, this result is independent of the initial conditions, in the sense that though originally (at the cathode) D might have been a function not only of u_0 but also of the direction, ultimately, over long intervals it can be a function of u only.

One can therefore apply the law of uniform distribution in the accessible part of phase space to both the principal cases under consideration, to the magnetron and to the space-charge electron lens, as shown in figure 1, though for different reasons. In the first case it is justified by Liouville's theorem, in the second by the ergodic theorem. The assumption of very long trajectories appears well justified in this second case, as the cathode can be made very small in comparison to the rest of the volume, which the electrons will traverse in general many times until they return to the cathode.

It may again be emphasized that the problem is to be treated dynamically as a *one-electron problem*, and that electron interaction must be expected to make itself particularly felt in the case of long trajectories. It may be repeated that the present purpose is not to establish a generally valid theory of magnetrons and similar devices, but to build a basis for such a theory by first investigating the conclusions of a theory without interaction.

3. THE LAW OF SPACE-CHARGE DISTRIBUTION

On the basis of this assumption the volume of the accessible momentum (or velocity) space associated with any point is a measure of the space-charge density at this point. To calculate this, one puts, as in figure 2,

$$u_t^2 = u^2\cos^2\alpha, \quad u_z^2+u_r^2 = u^2\sin^2\alpha. \tag{16}$$

---

* It appears that the present application is proof against the weighty arguments which have been advanced against considering the ergodic or the quasi-ergodic theorem as the basis of classical statistical mechanics. Cf. R. C. Tolman (1938), p. 70.

Substitution into equation (10) (with the equality sign) gives the following equation for the intersections of the sphere of radius u with the quadric $Q = 0$:

$$(u\cos\alpha - \mathscr{A})^2 = \left(\frac{a}{r}\right)^2 (u^2 - \phi) = \left(\frac{a}{r}u_0\right)^2. \tag{17}$$

Physically this means that the extreme electrons reaching a point will be those which started at the cathode in tangential direction. The limits are (u_0 being considered always as positive)

$$u\cos\alpha = \mathscr{A} \pm \frac{a}{r}u_0. \tag{17·1}$$

Four cases must now be distinguished, which give four different laws for the space-charge distribution:

(*a*) The equation (17) has two real roots, α_1, α_2. The condition for this is

$$\left(\mathscr{A} + \frac{a}{r}u_0\right)^2 < u^2 = \phi + u_0^2. \tag{a}$$

If this condition is fulfilled (as in figure 2), the volume of the shell segment is, apart from a factor 2π,

$$(\cos\alpha_1 - \cos\alpha_2)\,u^2 du.$$

Combined with (17·1), which gives

$$u(\cos\alpha_1 - \cos\alpha_2) = 2\frac{a}{r}u_0, \tag{17·2}$$

and taking into consideration that $u\,du = u_0\,du_0$, one obtains for the charge density, apart from a constant factor, which may be included in the phase density D,

$$d\rho = \frac{a}{r}D(u_0)\,u_0^2\,du_0. \tag{18}$$

This means that if electrons are emitted only in an infinitesimal energy interval ('microcanonic assembly') the density will be inversely proportional to the distance from the axis, independently of the electric and magnetic fields, which determine only the limits inside which it is valid, but do not interfere with the distribution itself. This strikingly simple law holds also if the primary emission is not homogeneous, up to a certain maximum initial velocity. This law is the most important, but for completeness the other cases will also be discussed.

(*b*) The equation (17) has only one real root. The condition for this is

$$\left(\mathscr{A} - \frac{a}{r}u_0\right)^2 < \phi + u_0^2 < \left(\mathscr{A} + \frac{a}{r}u_0\right)^2. \tag{b}$$

In this case the spherical segment becomes a spherical cap. The accessible volume is

$$(1 - \cos\alpha)\,u^2 du,$$

and the density will follow the law

$$d\rho = \left[\sqrt{(\phi+u_0^2)}-\mathscr{A}+\frac{a}{r}u_0\right]D(u_0)\,u_0^2du_0. \tag{19}$$

Unlike (a) this depends explicitly on ϕ and $\mathscr{A}$.

If equation (17) has no real roots, this can mean two things:

(c) The whole volume of the spherical shell is accessible. The law of density is

$$d\rho = \sqrt{(\phi+u_0^2)}\,D(u_0)\,u_0^2du_0. \tag{20}$$

This is the case if the whole sphere is outside the quadric, which in this case must be an ellipsoid.

(d) If, on the other hand, the whole sphere is inside the quadric, the accessible volume is zero, and so is the density. Therefore in the following this case will be referred to as (o).

One must now determine the range of validity of these laws, in terms of ϕ, $\mathscr{A}$, u_0 and r (or x). The discussion, though elementary, is rather complicated, and only the results will be given. The velocity ranges in which different solutions (a)–(d) are valid are separated by roots of the equation

$$\left(\mathscr{A}+\frac{a}{r}u_0\right)^2 = \phi+u_0^2,$$

which are
$$u_1,\, u_2 = \frac{x}{x^2-1}\left[\mathscr{A}\pm\sqrt{\{\mathscr{A}^2-(x^2-1)\,(\phi-\mathscr{A}^2)\}}\right]. \tag{21}$$

Of these roots, if they exist, the smaller one is called u_1 and the larger one u_2. It will also be convenient to introduce the abbreviation

$$\kappa = x^2/(x^2-1).$$

The following table gives the intervals of u_0 in which the laws (a), (b), (c) and (d) are valid:

$x<1$	$\phi>\mathscr{A}^2$	$\mathscr{A}^2>\phi>\kappa\mathscr{A}^2$	$\kappa\mathscr{A}^2>\phi$
(a)	$0-u_2$	—	—
(b)	—	u_1-u_2	—
(c)	$u_2-\infty$	$u_2-\infty$	—
(o)	—	$0-u_1$	$0-\infty$

$x>1$	$\phi>\kappa\mathscr{A}^2$	$\kappa\mathscr{A}^2>\phi>\mathscr{A}^2$	$\mathscr{A}^2>\phi$
(a)	$0-\infty$	$0-u_1,\ u_2-\infty$	$u_2-\infty$
(b)	—	u_1-u_2	u_1-u_2
(c)	—	—	—
(o)	—	—	$0-u_1$

This table is illustrated in figure 4 in the particular case $a = r_0$, i.e. if the cathode is in the homogeneous part of the field. In this case the curves which divide the field are

$$\phi/K^2 = \left(x-\frac{1}{x}\right)^2, \quad \phi/K^2 = x^2-1,$$

as given by equation (13). In the general case the curve $\phi = \kappa\mathscr{A}^2$ has a different shape, as it goes to $\pm\infty$ at $x = 1$, but the relative position of the two curves remains the same, and nothing essential is changed.

The table discloses the extraordinary variety of equilibrium shapes of rotating electron swarms which could be obtained with large initial velocities. It may be somewhat surprising that the initial velocity adds to the manifoldness of the problem, as one might think that lowering of the cathode potential would be equivalent to an increase of u_0. But u_0 appears in the equations not only in the combination $\phi + u_0^2$, but also in the combination au_0/r. This is a consequence of the momentum integral, and means physically that a velocity u_0 at the radius a is 'worth' a velocity au_0/r at the radius r. Things become simple only at small initial velocities, as in this range there obtains either the simple hyperbolical law (*a*) or the law (*o*).

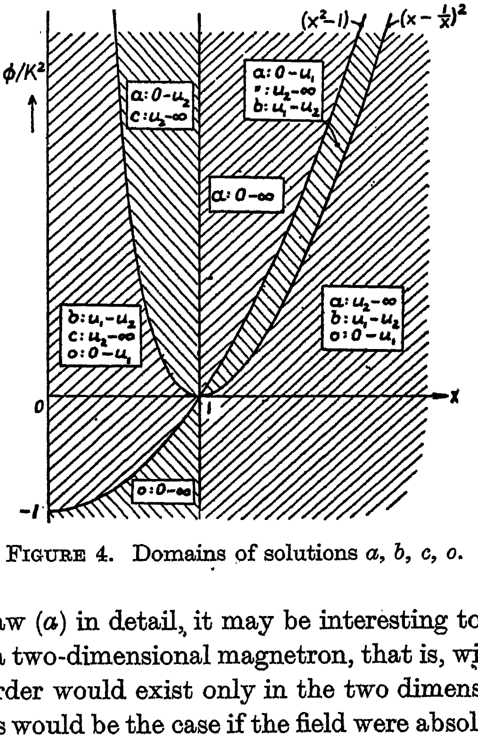


FIGURE 4. Domains of solutions *a*, *b*, *c*, *o*.

Before discussing law (*a*) in detail, it may be interesting to compare it with the corresponding law in a two-dimensional magnetron, that is, with an electron swarm in which ergodic disorder would exist only in the two dimensions r and θ, but not in the z-direction. This would be the case if the field were absolutely cylindrical, and if the electrons had definite, e.g. zero, initial components u_{z0}. In this case the phase volume changes into an area, which is

$$(\alpha_2-\alpha_1)\,u\,du = \left[\cos^{-1}\frac{\mathscr{A}-au_0/r}{\sqrt{(\phi+u_0^2)}} - \cos^{-1}\frac{\mathscr{A}+au_0/r}{\sqrt{(\phi+u_0^2)}}\right]u\,du, \tag{22}$$

and series development gives the approximate formula for the density

$$\rho \sim (\alpha_1-\alpha_2) \cong \frac{1}{\mathscr{A}+au_0/r}\left(\sqrt{[\phi-\mathscr{A}^2+2au_0\mathscr{A}/r]}-\sqrt{[\phi-\mathscr{A}^2-2au_0\mathscr{A}/r]}\right). \tag{22·1}$$

If $\phi - \mathscr{A}^2 \gg (2a/r)\, u_0 \mathscr{A}$, this can be further simplified and gives

$$\rho = \text{const.}/r\sqrt{(\phi - \mathscr{A}^2)}. \tag{22·2}$$

This is a law entirely different from any of the previous ones. Such sudden changes of law with the number of degrees of freedom are familiar in classical statistical mechanics, and are a warning to be cautious in the application of statistical principles. But in the present case there is good reason to think that the axial degree of freedom can be considered as 'fully excited'. In the case of the magnetron with infinite cylindrical cathode this follows without any application of statistics from Liouville's theorem, and the arrangement in figure 1 cannot be realized without departing from cylindricity.

It may be mentioned that a further step in the same direction, i.e. restricting the disorder of the system, would lead to Hull's solution, with the corrections applied by Brillouin.

4. SELF-CONSISTENT STATIONARY ELECTRON SWARMS

The main result of the foregoing discussion is that for very small initial velocities the density will follow the simple law $1/r$ in the region in which $\phi > \mathscr{A}^2$, and will be zero outside it. The more complicated laws (b) and (c) come into action only where $\phi \cong \mathscr{A}^2$. Hence, apart from two sheaths near the edges of the electron cloud, the $1/r$ law will obtain not only for electrons emitted in an infinitesimal energy interval, but also for electrons emitted by thermionic cathodes.

In order to obtain the potential distribution one must substitute this law into Poisson's equation

$$\frac{\partial^2\phi}{\partial z^2} + \frac{\partial^2\phi}{\partial r^2} + \frac{1}{r}\frac{\partial\phi}{\partial r} = \frac{C}{r}, \tag{23}$$

where C is a constant, to be determined later. The solution will be discussed only in the cylindrical case $\partial^2\phi/\partial z^2 = 0$, which applies also approximately to arrangements like the one in figure 1, at some distance from the ends where the magnetic intensity falls off. Here the electron cloud can be prevented from escaping by two sufficiently negative end-plates.

The general solution in the cylindrical case is

$$\phi = Cr + c_1 + c_2 \log r/r_0, \tag{24}$$

where c_1 and c_2 are constants, and r_0 is as defined in figure 1.

The electron cloud has, in the present approximation, sharp boundaries at $r = r_1$ inside and $r = r_2$ outside. At both edges the velocity is purely tangential. The boundary conditions are

$$\phi(r_1) = \mathscr{A}^2(r_1), \quad \phi(r_2) = \mathscr{A}^2(r_2), \tag{25}$$

and in addition at the inner edge $\quad d\phi/dr = 0, \qquad (26)$

if there is no charge inside r_1. This condition is valid in the case of figure 1 because there is no electrode at the axis, but it is valid also in the case of the cylindrical magnetron, as the field strength at the cathode must be zero in the equilibrium condition. There are therefore three conditions for the five constants C, c_1, c_2, r_1 and r_2, and two degrees of freedom are left over. One of these corresponds to the anode potential, which can have any value below the critical potential, and the other to the degree of saturation. The total charge in the cloud may assume any value up to a certain maximum. This will be reached if $r_1 = r_0$ and $\phi(r_0) = 0$, which means that no further electrons can reach the cloud from the cathode.

Having assumed ϕ independent of z, the same assumption must now be made for $\mathscr{A}$. Hence, neglecting the magnetic effect of the rotating electrons, which will be dealt with later, a homogeneous magnetic field must be assumed, writing

$$\mathscr{A}^2 = \frac{eH^2}{8mc^2} r_0^2 \left(\frac{r}{r_0} - \frac{r_0}{r}\right)^2. \tag{27}$$

Substituting this into the boundary conditions (25) and (26), the following expression is obtained for the space-charge density at a radius r:

$$\rho(r) = \tfrac{1}{4}\rho_H \frac{(r_2^2 - r_1^2)\,[1-(r_0^2/r_1 r_2)^2]}{r[r_2 - r_1(1+\log r_2/r_1)]}, \tag{28}$$

where ρ_H is Hull's space-charge density, as defined in equation (1). This is the solution of the problem, with two free parameters r_1 and r_2. It is illustrated in figure 5 in the special case $r_2/r_0 = 3$ for various radii r_1, which correspond to various degrees of saturation. The total charge contained between r_1 and r_2 is marked on the curves in percentages of the maximum charge.

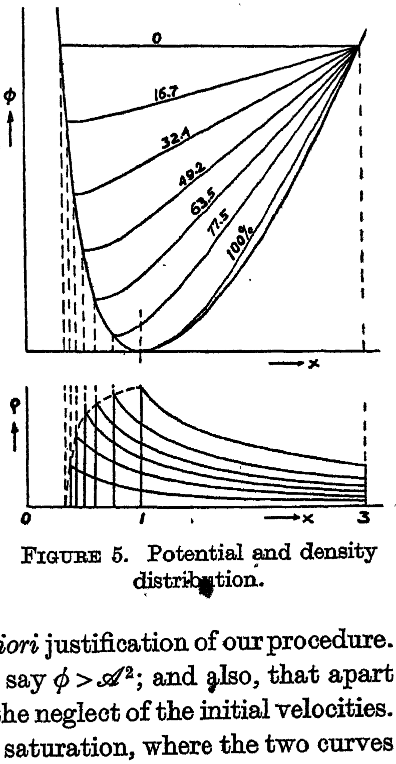


FIGURE 5. Potential and density distribution.

It may be noted that this result gives an *a posteriori* justification of our procedure. The law (a) has been assumed to hold, that is to say $\phi > \mathscr{A}^2$; and also, that apart from the edges ϕ exceeds $\mathscr{A}^2$ sufficiently to justify the neglect of the initial velocities. This assumption is now verified, except at 100 % saturation, where the two curves come rather close together, near the inner edge r_0.

In the case of maximum saturation, i.e. $r_1 = r_0$, the density at the outer edge becomes, according to equation (28),

$$\rho(r_2) = \tfrac{1}{4}\rho_H \frac{[1-(r_0/r_2)^2]^2}{1-(r_0/r_2)\log(r_2/r_0)}, \tag{28·1}$$

and for large ratios r_2/r_0 this approaches *one-quarter of Hull's value*. Applying this to the cut-off characteristic of the cylindrical magnetron an interesting difference arises between the present and previous theories. The sudden jump of current when the critical voltage is exceeded ought to be only about one-quarter of the saturation value, instead of the whole amount. There are, indeed, measured magnetron characteristics which show this phenomenon,* but it would be premature to consider this as a confirmation of the theory. The present theory cannot account—and no one-electron theory possibly could account—for the discrepancy of the voltage at which current starts to flow. Also, the present theory is valid only so long as the cloud is static, i.e. so long as there is no current. There is the possibility that even small currents might modify the calculated distribution very considerably.

The density at the inner edge r_1 in the case of saturation is about $\frac{1}{4}\rho_H r_2/r_0$, which can far exceed ρ_H. But this is not necessarily the largest density which can be maintained in a magnetron of prescribed outer radius. For very large ratios r_2/r_1, the density at the inner edge may be written approximately

$$\rho(r_1) \simeq \tfrac{1}{4}\rho_H \frac{r_2}{r_1}\left[1-\left(\frac{r_0^2}{r_1 r_2}\right)^2\right] = \tfrac{1}{4}\rho_H \left(\frac{r_2}{r_0}\right)^2 \left(\frac{r_0^2}{r_1 r_2}\right)\left[1-\left(\frac{r_0^2}{r_1 r_2}\right)^2\right]. \tag{28·2}$$

This has a maximum at $$r_1 = \sqrt{(3)}\, r_0^2/r_2,$$

i.e. when r_1 is 1·73 times larger than the smallest radius which electrons can reach at all at a given outer radius. The value of the maximum is

$$\rho(r_1)_{\text{max.}} = 0{\cdot}094\rho_H (r_2/r_0)^2, \tag{28·3}$$

which can exceed the value of $\rho(r_1)$ at maximum saturation by any amount if r_2/r_0 is made sufficiently large. This effect begins to develop at about $r_2/r_0 = 4$. In the example shown in figure 5 the largest density is still reached at maximum saturation.

In what limits would such an electron cloud act as a dispersing lens? The radial electric gradient is

$$\frac{d\phi}{dr} = C + c_2/r = C(1 - r_1/r) = -4\pi\rho(r_1)\, r_1(1 - r_1/r). \tag{29}$$

This must be compared with the gradient in the case of Hull's distribution, which would just suffice to keep the electrons in equilibrium

$$\left(\frac{d\phi}{dr}\right)_H = -2\pi\rho_H r.$$

Substituting in equation (29) the maximum value of $\rho(r_1)$ from equation (28·3), the following criterion is obtained for a dispersing lens:

$$\frac{d\phi/dr}{(d\phi/dr)_H} = 0{\cdot}188\left(\frac{r_2}{r_1}\right)^2 \frac{r_1}{r}\left(1 - \frac{r_1}{r}\right) > 1. \tag{30}$$

A dispersing effect will therefore always exist in a certain interval of r so long as $r_2/r_1 > 4{\cdot}6$. But this effect is confined between two limiting radii, between which it reaches a maximum value.

---

* Cf. Harvey (1943, p. 105, figure 43*a*).

This is not the law required of a dispersing lens, even less of a lens which could be used to correct the aberrations of ordinary electron lenses. In a correcting lens the density is required to start at the axis with a value exceeding Hull's value, and to increase approximately parabolically with the distance from it.

It appears likely that a distribution similar to the required one would establish itself in any arrangement of the type as shown in figure 1, unless it is constructed with extraordinary precision. At some distance from the cathode, where the magnetic field has become appreciably homogeneous, the 'axis' of the magnetic field loses its significance, and the momentum integral is valid only if the electric field is rigorously rotationally symmetric and perfectly alined with the cathode. Very small departures from these ideal conditions will cause the charge distribution to depart appreciably from the calculated form. The hollow tube inside the radius r_1 will fill up with electrons, and one may expect that within certain limits the real distribution will approach the desired shape.

5. EXTENDED ELECTRON SOURCE

These processes may be followed to a certain point by a simple extension of the foregoing theory, if one considers cathodes of a certain radial extension instead of filament cathodes with vanishing thickness and perfect alinement.

The density contributed by the emission of a cathode of vanishing thickness, which emits in the limits u_0, u_0+du_0, was found to be

$$d\rho = \frac{a}{r} D(u_0)\, u_0^2 du_0. \tag{18}$$

Now assume that u_0 and du_0 go to zero, whilst the product $D(u_0)\, u_0^2 du_0$ divided by the radial extension da of the cathode approaches a finite limit. It will be convenient to express a by r_0 (as defined in figure 1) and include it in a function $F(r_0)$, with which the density due to the emission of an infinitesimal strip of the cathode is written

$$d\rho = \frac{1}{r} F(r_0)\, dr_0. \tag{31}$$

If the cathode has constant emission density, the function $F(r_0)$ is proportional to the cathode area corresponding to the limits r_0, r_0+dr_0, multiplied with the radius a.

Assume in the following, to simplify matters, that the cathode extends to $r_0 = 0$, i.e. to the flux line $\psi = 0$. Investigating only a region near the axis in which only a part of the cathode is active, the area outside a maximum radius $r_{0m.}$ is cut off from contributing to the density by the momentum integral. This gives the following law for the potential and density distribution:

$$-4\pi\rho = \frac{d^2\phi}{dr^2} + \frac{1}{r}\frac{d\phi}{dr} = \frac{C}{r}\int_0^{r_{0m.}} F(r_0)\, dr_0. \tag{32}$$

The radius $r_{\text{om.}}$ is determined by $\phi = \mathscr{A}^2$, and in a homogeneous magnetic field (equation (27)) this gives

$$r^2_{\text{om.}} = r^2 + \frac{r}{K}\sqrt{\phi}. \tag{33}$$

As a particularly simple example assume $F(r_0) = \text{const.}$ This gives a density proportional to

$$\sqrt{\left(1+\frac{1}{Kr}\sqrt{\phi}\right)}.$$

Putting $\phi \sim r^2$ this becomes a constant, and the equation (32) is satisfied. Hence, in order to realize a dispersing lens in which at least near the axis the radial force is exactly proportional to the radius, the cathode must be arranged in such a way as to make $F(r_0) = \text{const.}$ This problem can easily be solved once the magnetic field is given. How far such a lens would in fact possess the desired qualities, and how far electron interaction would interfere with its performance, only experiment can show.

6. MAGNETIC EFFECT OF THE ROTATING SPACE CHARGE

Up to now, in the examples though not in the general formulae, a homogeneous magnetic field has been assumed. One must now check under what conditions this assumption is justified.

The rotating electron swarm represents a ring current of intensity $i_t = \rho\bar{v}_t$. By a well-known proposition regarding the centre of gravity of a spherical shell segment

$$\bar{v}_t = \tfrac{1}{2}v(\cos\alpha_1 + \cos\alpha_2),$$

and introducing again u instead of v, from equation (17·1) the simple result

$$\bar{u}_t = \mathscr{A} \tag{34}$$

is obtained.

Outside the radius r_0 the cloud rotates in the direction of the vector potential A, inside it in the opposite direction. It may be noted that the maximum departures from the average $\mathscr{A}$ are $\pm au_0/r$. Hence on approaching the axis one finds increasing tangential velocity differences, at the same time as increasing electron density, so that electron interaction ought to play a particularly prominent part in this region.

The ring current has such a sign that outside r_0 it opposes the current which produces the magnetic field, whereas inside r_0 it increases the field. For large ratios r_2/r_0 the shielding effect far predominates.

Consider now only the middle region of the magnetron, in which the magnetic field is parallel to the z-direction. The law of its distribution is

$$-\operatorname{curl} H = \frac{dH_z}{dr} = -\frac{4\pi}{c}i_t = -\frac{4\pi}{c}\rho v_t. \tag{35}$$

Using equations (3) and (34) H_z and v_t is expressed by means of the vector potential A, and denoting differentiation with respect to r by primes,

$$A''+\frac{1}{r}A'-\frac{1}{r^2}A=\frac{1}{\lambda^2}\left[A-\frac{a}{r}A_0\right]\left(\frac{\rho}{\rho_H}\right), \tag{36}$$

is obtained, where

$$\lambda^2=2\frac{mc^2}{eH}=\frac{1}{2}\left(\frac{c}{\omega_H}\right)^2. \tag{37}$$

λ is a characteristic length, which is $1/2\pi\sqrt{2}=0{\cdot}1125$ of the vacuum wave-length associated with the frequency ω_H. Numerically

$$\lambda H=2410\,\text{cm.gauss}. \tag{37·1}$$

The effect of the ring current will be appreciable only in large magnetrons in which the radius is of the order of λ.

Now write the charge density in the form

$$\rho=\rho_H\frac{\rho(r_2)}{\rho_H}\frac{r_2}{r}=\rho_H\frac{R}{r}. \tag{38}$$

The characteristic radius R will approach $\frac{1}{4}r_2$ for large ratios r_2/r_0. The equation (36) now becomes

$$A''+\frac{1}{r}A'-\left(\frac{1}{r^2}+\frac{R}{\lambda^2 r}\right)A=-\frac{Ra}{(\lambda r)^2}A_0. \tag{36·1}$$

This has the particular solution $A=aA_0/r$,

which is of no importance, as it does not contribute to the magnetic field. The solution of the homogeneous equation is

$$A=Z_2(i\sqrt{[4Rr/\lambda^2]}), \tag{39}$$

i.e. a cylindrical function of the second order with imaginary argument. The solution is therefore a sum, with constant coefficients of the modified Bessel and Hankel functions $I_2(\xi)$ and $K_2(\xi)$, if the symbol ξ is introduced for the dimensionless parameter $\sqrt{(4Rr/\lambda^2)}$.

For the constants there are two boundary conditions. The first is that at the outer radius r_2 the field intensity H_z must assume a prescribed value. Substituting equation (39) into equation (3) and using the well-known relation between cylindrical functions

$$Z_n'+\frac{n}{\xi}Z_n=Z_{n-1},$$

then

$$\xi_2[C_1I_1(\xi_2)+C_2K_1(\xi_2)]=2r_2H(r_2), \tag{40}$$

where ξ_2 is the value of ξ corresponding to r_2.

The second boundary condition is not so obvious, as the field intensity at the inner radius r_1 is unknown. The condition is obtained by the following consideration. Though the solution does not extend to the axis, but only to the radius r_1, it ought to be possible to continue it, by adding ring currents inside, according to the same law, without modifying the solution outside r_1, as currents inside a radius have no effect on the field outside it. This process of continuation leads at the axis to a magnetic field of the same direction as the outer field, which goes to infinity like $1/r$. This gives the condition from which the coefficient C_2 of the modified Hankel function K_2 can be determined.

In the following, to simplify the discussion, only the case $A_0 = 0$ will be considered, i.e. the cathode arranged on the line $\psi = 0$, in which case the electrons arrive at the axis with zero tangential velocity and no singularity arises. In this case the whole mass of electrons rotates in one direction, so as to oppose the outer field by its magnetic effect. The solution in this case is given by the modified Bessel function $I_2(\xi)$, and the magnetic intensity follows the law

$$\frac{H(r)}{H(r_2)} = \sqrt{\left(\frac{r_2}{r}\right)} \frac{I_1(\sqrt{[4Rr/\lambda^2]})}{I_1(\sqrt{[4Rr_2/\lambda^2]})}. \tag{41}$$

For small arguments $I_1(\xi)/\xi$ approaches the limit $\frac{1}{2}$. Hence the magnetic field intensity at the axis becomes, if H is written for $H(r_2)$,

$$H_0 = \tfrac{1}{2}H\xi_2/I_1(\xi_2), \tag{41·1}$$

and as in this case $R = \frac{1}{4}r_2$, this can be written

$$H_0 = \tfrac{1}{2}H(r_2/\lambda)/I_1(r_2/\lambda). \tag{41·2}$$

The following table, calculated from the data of Jahnke & Emde (*Funktionentafeln*), shows the shielding effect of the ring current for magnetrons of different radius, measured in units of λ:

r_2/λ	0·25	0·5	1	2	3	4
H_0/H %	99·3	97·0	88·5	62·9	38·9	20·5
r_2/λ	5	6	7	8	9	10
H_0/H %	10·2	4·78	2·24	1·00	0·437	0·187

The effect becomes therefore very strong in large magnetrons, not too large to be practicable. With $H = 5000$ gauss, for instance, λ is a little less than 0·5 cm.; hence with a radius of 5 cm. it ought to be possible to reduce the magnetic field at the axis to less than 100 gauss. According to the present theory, therefore, it ought to be possible to produce extraordinary concentrations of free electrons, far larger than ever produced experimentally, and to study them under favourable conditions in a relatively weak magnetic field. It may well be possible to produce concentrations sufficient for observable optical effects. But it must be borne in mind that the higher the electron density the stronger the interaction, and the more remote the present theory must be from reality.

I am indebted to Dr C. J. Milner for many discussions on this subject. I thank the Directors of the British Thomson Houston Company for permission to publish this paper.

REFERENCES

Becker, R. 1933 *Theorie der Elektrizität.* 2. *Elektronentheorie.* Leipzig: Teubner.
Brillouin, L. 1941 *Phys. Rev.* **60**, 385.
Brillouin, L. 1942 *Phys. Rev.* **62**, 166.
Harvey, A. F. 1943 *High frequency thermionic tubes.* London: Chapman and Hall.
Hull, A. W. 1921 *Phys. Rev.* **18**, 13.
Hull, A. W. 1924 *Phys. Rev.* **23**, 112.
Langmuir, I. 1927 *Z. Phys.* **46**, 271.
Langmuir, I. 1928 *Proc. Nat. Acad. Sci., Wash.*, **14**, 627.
Linder, E. G. 1938*a* *Proc. Inst. Radio Engrs, N.Y.*, **26**, 346.
Linder, E. G. 1938*b* *J. Appl. Phys.* **9**, 331.
Pidduck, F. B. 1936 *Quart. J. Math.* **7**, 201.
Schwarzschild, K. 1903 *Nachr. Ges. Wiss. Göttingen*, **3**, 126.
Tolman, R. C. 1938 *The principles of statistical mechanics.* Oxford Univ. Press.

The electrical conductivity of an ionized gas in a magnetic field, with applications to the solar atmosphere and the ionosphere

By T. G. Cowling, *The University, Manchester*

(*Communicated by S. Chapman, F.R.S.—Received* 26 *July* 1944)

The methods of Chapman and Enskog are used to discuss conduction of electricity and diffusion currents in an ionized gas with several constituents, in a transverse magnetic field. The free-path formula for the conductivity is compared with that derived by the exact methods. The two formulae are identical in form if a correction is applied to the usual free-path method; this correction robs the method of much of its simplicity. The uncorrected free-path method, however, gives correct results for the electron contribution to the conductivity in all practical cases; and for the ion contribution if a large number of neutral molecules are present—e.g. in the earth's upper atmosphere, about 5×10^5 times the number of ions (of both signs).

Numerical values are given for the conductivity in the sun's outer layers and in the earth's upper atmosphere. Mechanical forces due to currents induced in moving material are shown to be very important in the sun, and in the F-layer of the earth's atmosphere. The solar results are used to discuss the motion of solar prominences and eruptions. In the earth's atmosphere, the observed collision frequencies of electrons are shown to imply upper limits for ion-densities in the E and F layers. The integral conductivities of the E and F layers are estimated, and it is shown that, on the dynamo theory of the lunar variation of the earth's magnetic field, tidal oscillations in these layers must be between 100 and 1000 times as great as those at the ground. Diamagnetism and drift currents are shown to make negligible contributions to the lunar and solar variations of the earth's magnetic field.

In an Appendix, the applicability of Boltzmann's equation to strongly ionized gases is discussed.

1. INTRODUCTION

In discussing diffusion, and the conduction of electricity, in an ionized gas in a transverse magnetic field, some authors have used the free-path method, and others the more exact 'velocity-distribution' method, originally developed by Chapman and Enskog. There are certain discrepancies between the results of the two methods; because of the relative simplicity of the free-path method, it has often been preferred to the other.

In the present paper, the conductivity is further discussed. First, the discrepancies between results of the two methods are shown to be due to the assumptions implicit in them, and to an internal inconsistency in the free-path method, as often used. Thereafter, the velocity-distribution method alone is used. Examination is made of the error involved in using first approximations to the velocities of diffusion in a binary gas. Similar first approximations are then used to discuss the conductivity of a gas with more than two constituents. The results are found, in certain important special cases, to resemble those derived by free-path methods. Electric currents arising from gradients of composition and pressure are also discussed. Applications to the atmospheres of the sun and the earth are briefly considered.

2. EARLIER RESULTS

Suppose that a number of types of particle (neutral molecules, atoms, ions and electrons) are present; these are regarded as molecules of distinct gases. The mass, charge (in e.m.u.) and velocity of a molecule of the rth gas are denoted by m_r, e_r, $\boldsymbol{c}_r$; the number-density, mass-density, partial pressure and velocity-distribution function for the rth gas are n_r, ρ_r, p_r and f_r; the number-density, mass-density, pressure and temperature of the gas as a whole are n, ρ, p and T. The mass-velocity $\boldsymbol{c}_0$ of the gas is defined by

$$\rho \boldsymbol{c}_0 = \sum_r \rho_r \overline{\boldsymbol{c}}_r, \tag{1}$$

where $\overline{\boldsymbol{c}}_r$ is the mean velocity of molecules of the rth gas; and the peculiar velocity $\boldsymbol{C}_r$ of a molecule is defined by

$$\boldsymbol{C}_r = \boldsymbol{c}_r - \boldsymbol{c}_0.$$

The gas as a whole is supposed to be electrically neutral; thus

$$\sum_r n_r e_r = 0. \tag{2}$$

The electric current-density $\boldsymbol{j}$ can be expressed in the form

$$\begin{aligned} \boldsymbol{j} &= \sum_r n_r e_r \overline{\boldsymbol{c}}_r = \sum_r n_r e_r (\boldsymbol{c}_0 + \overline{\boldsymbol{C}}_r) \\ &= \sum_r n_r e_r \overline{\boldsymbol{C}}_r. \end{aligned} \tag{3}$$

It therefore does not depend on the mass-velocity, but only on the velocities of diffusion of the different gases through the mass.

The force on a molecule is divided into two parts. One part, $e_r \boldsymbol{C}_r \wedge \boldsymbol{H}$, arises from the peculiar velocity $\boldsymbol{C}_r$ of the molecule in the magnetic field $\boldsymbol{H}$. The other part, $m_r \boldsymbol{F}_r$, is independent of the peculiar velocity; it may include a term $e_r \boldsymbol{c}_0 \wedge \boldsymbol{H}$ due to the mass motion of the gas. Since diffusion parallel to a magnetic field is known to be unaffected by that field, we consider only the case in which $\boldsymbol{F}_r$ and all other factors producing diffusion are perpendicular to $\boldsymbol{H}$. Thermal diffusion is not considered; that is, temperature gradients are supposed absent.

A simple form of the free-path theory introduces the 'spiral-frequency' $\omega_r = e_r H/m_r$, and the mean time τ_r between successive collisions of a molecule of the rth gas with molecules of other gases. The collision-interval τ_r is supposed not to depend on $\boldsymbol{c}_r$. It is therefore to be regarded as a mean collision-interval effective in diffusion; a comparison with velocity-distribution theory shows that, for rigid elastic spherical molecules, it must be taken as three-quarters of the collision-interval defined by the ratio of the mean free path to the mean peculiar speed. The free-path theory expresses the velocity of diffusion $\overline{\boldsymbol{C}}_r$ of the rth gas by

$$\overline{\boldsymbol{C}}_r = \left\{\tau_r\left(\boldsymbol{F}_r - \frac{1}{\rho_r}\frac{\partial p_r}{\partial \boldsymbol{r}}\right) - \frac{\omega_r}{H}\tau_r^2\boldsymbol{H}\wedge\left(\boldsymbol{F}_r - \frac{1}{\rho_r}\frac{\partial p_r}{\partial \boldsymbol{r}}\right)\right\}\Big/(1+\omega_r^2\tau_r^2) \tag{4}$$

(Chapman and Cowling 1939, p. 327). This implies that the diffusion of the rth gas is due to

$$\boldsymbol{F}_r - \frac{1}{\rho_r}\frac{\partial p_r}{\partial \boldsymbol{r}}, \tag{5}$$

i.e. to factors depending solely on the rth gas. It has two components, one parallel to the vector (5), and one perpendicular to both this and $\boldsymbol{H}$; these are the 'direct' and 'transverse' velocities of diffusion. The electric current-density, given by (3), is the sum of contributions arising independently from the different gases.

The discussion by the velocity-distribution method has hitherto been limited to a binary mixture. It gives, as a first approximation to the velocity of relative diffusion of the constituent gases,

$$\overline{\boldsymbol{C}}_1 - \overline{\boldsymbol{C}}_2 = -\frac{p\rho}{\rho_1\rho_2}\,\frac{(\tau\boldsymbol{d}_{12} - H^{-1}\omega\tau^2\boldsymbol{H}\wedge\boldsymbol{d}_{12})}{1+\omega^2\tau^2} \tag{6}$$

(Chapman & Cowling 1939, p. 335), where

$$\omega = H(e_1 m_2\rho_2 + e_2 m_1\rho_1)/\rho m_1 m_2 = (\omega_1\rho_2 + \omega_2\rho_1)/\rho, \tag{7}$$

$$\tau = (\tau_1\rho_2 + \tau_2\rho_1)/\rho, \tag{8}$$

and

$$\boldsymbol{d}_{12} = -\frac{\rho_1\rho_2}{p\rho}\left\{\boldsymbol{F}_1 - \frac{1}{\rho_1}\frac{\partial p_1}{\partial \boldsymbol{r}} - \boldsymbol{F}_2 + \frac{1}{\rho_2}\frac{\partial p_2}{\partial \boldsymbol{r}}\right\}$$

$$= \frac{\partial}{\partial \boldsymbol{r}}\left(\frac{n_1}{n}\right) + \frac{n_1 n_2(m_2 - m_1)}{np\rho}\frac{\partial p}{\partial \boldsymbol{r}} - \frac{\rho_1\rho_2}{p\rho}(\boldsymbol{F}_1 - \boldsymbol{F}_2). \tag{9}$$

Since, from (1),

$$\sum_r \rho_r\overline{\boldsymbol{C}}_r = 0, \tag{10}$$

equation (6) determines the velocities of diffusion $\overline{\boldsymbol{C}}_1$, $\overline{\boldsymbol{C}}_2$ of the two constituent gases.

Equations (4) and (6) have a superficial similarity, but this hides important differences. Equation (4), for a binary gas, makes the velocities of diffusion $\overline{C}_1$, $\overline{C}_2$ depend separately on the vectors

$$F_1 - \frac{1}{\rho_1}\frac{\partial p_1}{\partial r}, \quad F_2 - \frac{1}{\rho_2}\frac{\partial p_2}{\partial r};$$

equation (6) implies that each depends on the difference between the vectors. Again, (4) makes $\overline{C}_1$, $\overline{C}_2$ depend respectively on ω_1, τ_1 and ω_2, τ_2; (6) shows that each depends on ω, τ, which are weighted means of ω_1, ω_2 and τ_1, τ_2.

For example, if $n_1 = n_2$, $m_1 = m_2$, and $e_1 = -e_2$, (7) gives $\omega = 0$; thus (6) indicates no transverse diffusion, and a direct diffusion unaffected by the magnetic field. In this case, in fact, particles of the two gases have accelerations equal in magnitude and direction by reason of their equal and opposite velocities of diffusion in the magnetic field. Thus the field produces an acceleration of the gas as a whole; the diffusion of one constituent relative to the other is affected by the field only because the change in mass-velocity reacts on the forces producing diffusion. In indicating a transverse diffusion and a reduction in the direct diffusion, free-path theory confuses mass-motion effects with diffusion.

Free-path theory in any case tacitly assumes that the gas is not accelerating. It calculates the mean velocity of molecules at a given point by grouping them according to their velocities at the beginning of the current free path, and in doing so it makes no distinction between the values of the mass-velocity at the beginning of the free path and at the instant considered. This tacit assumption must be borne in mind in comparisons with the velocity-distribution theory, which allows for an acceleration, though assuming that it produces only small changes in mass-velocity during a free path. But the results of the two theories are not identical even for an unaccelerated gas.

3. Free-path theory

The discrepancy between the results of the two theories arises because, in the free-path discussion leading to (4), it is assumed that, directly after a collision, a molecule has, on an average, no peculiar velocity, i.e. that its velocity is, on an average, c_0. This assumption cannot, of course, apply to collisions between like molecules, as such collisions cannot destroy the common velocity of diffusion; collisions between like molecules are, in fact, neglected in considering diffusion. But the assumption is also invalid for collisions between unlike molecules; if the velocities of molecules m_1, m_2 just after a mutual collision are distributed at random about a mean value c', then (remembering that the collision-interval is supposed independent of the molecular velocity) c' must satisfy the momentum equation

$$(m_1+m_2)\,c' = m_1\bar{c}_1 + m_2\bar{c}_2 \tag{11}$$

and so cannot in general equal c_0.

Consider an unaccelerated binary gas. It may, without loss in generality, supposedly be at rest, since a change in $\mathbf{c}_0$ is simply equivalent to a change in $\mathbf{F}_r$. The free-path argument yielding (4) assumed the mean velocity of molecules m_1, m_2 immediately after collision to be zero; assume now that it is $\mathbf{c}'$, given by (11). Then, the velocity of diffusion $\bar{\mathbf{C}}_r$ is increased by the velocity after a mean time τ_r of a molecule moving in the field $\mathbf{H}$ with initial velocity $\mathbf{c}'$, i.e. by

$$(\mathbf{c}' - H^{-1}\omega_r\tau_r \mathbf{H}\wedge\mathbf{c}')/(1+\omega_r^2\tau_r^2).$$

Hence*
$$\bar{\mathbf{C}}_r = (1+\omega_r^2\tau_r^2)^{-1}\left\{\mathbf{c}' + \mathbf{F}_r\tau_r - \frac{\tau_r}{\rho_r}\frac{\partial p_r}{\partial \mathbf{r}} - \frac{\omega_r\tau_r}{H}\mathbf{H}\wedge\left(\mathbf{c}' + \mathbf{F}_r\tau_r - \frac{\tau_r}{\rho_r}\frac{\partial p_r}{\partial \mathbf{r}}\right)\right\}, \tag{12}$$

and here $\bar{\mathbf{C}}_r = \bar{\mathbf{c}}_r$. Since $\mathbf{c}'$ is comparable with the diffusion velocity, the new terms in this equation are not unimportant.

Equation (12) gives
$$\bar{\mathbf{c}}_r - \frac{\omega_r\tau_r}{H}\bar{\mathbf{c}}_r\wedge\mathbf{H} = \mathbf{c}' + \mathbf{F}_r\tau_r - \frac{\tau_r}{\rho_r}\frac{\partial p_r}{\partial \mathbf{r}}.$$

Now $n_1/\tau_1 = n_2/\tau_2$, since each gives the frequency of collisions between unlike molecules per unit volume. Hence, using (8),

$$n_r/\tau_r = n_1 n_2(m_1+m_2)/\rho\tau,$$

and so
$$\frac{n_1 n_2(m_1+m_2)}{n_r\rho\tau}(\bar{\mathbf{c}}_r - \mathbf{c}') - \frac{\omega_r}{H}\bar{\mathbf{c}}_r\wedge\mathbf{H} = \mathbf{F}_r - \frac{1}{\rho_r}\frac{\partial p_r}{\partial \mathbf{r}}. \tag{13}$$

Multiply this equation by ρ_r and sum over values of r; then since $\rho_r\omega_r = n_r e_r H$, using (3) and (11),

$$\sum_r \rho_r\mathbf{F}_r + \mathbf{j}\wedge\mathbf{H} = \frac{\partial p}{\partial \mathbf{r}}.$$

This is the condition for zero acceleration of the gas, which, accordingly, follows as a direct consequence from (11).

Again, by (9) and (13),

$$\frac{\rho p}{\rho_1\rho_2}\mathbf{d}_{12} = \frac{m_1+m_2}{\rho\tau}\{n_1(\bar{\mathbf{c}}_2 - \mathbf{c}') - n_2(\bar{\mathbf{c}}_1 - \mathbf{c}')\} - H^{-1}(\omega_2\bar{\mathbf{c}}_2 - \omega_1\bar{\mathbf{c}}_1)\wedge\mathbf{H}.$$

But, using (11) and (1), with $\mathbf{c}_0 = 0$,

$$\rho_1\bar{\mathbf{c}}_1 = -\rho_2\bar{\mathbf{c}}_2 = \frac{n_1 n_2(m_1+m_2)}{n_2 - n_1}\mathbf{c}' = \frac{\rho_1\rho_2}{\rho}(\bar{\mathbf{c}}_1 - \bar{\mathbf{c}}_2).$$

Substituting these values for $\bar{\mathbf{c}}_1$, $\bar{\mathbf{c}}_2$, $\mathbf{c}'$ in terms of $\bar{\mathbf{c}}_1 - \bar{\mathbf{c}}_2$,

$$\begin{aligned}\frac{p\rho\tau}{\rho_1\rho_2}\mathbf{d}_{12} &= -(\bar{\mathbf{c}}_1 - \bar{\mathbf{c}}_2) + \frac{\tau}{\rho H}(\omega_1\rho_2 + \omega_2\rho_1)(\bar{\mathbf{c}}_1 - \bar{\mathbf{c}}_2)\wedge\mathbf{H}, \\ &= -(\bar{\mathbf{c}}_1 - \bar{\mathbf{c}}_2) + \frac{\omega\tau}{H}(\bar{\mathbf{c}}_1 - \bar{\mathbf{c}}_2)\wedge\mathbf{H}\end{aligned}$$

---

* This result can also be derived directly, using methods similar to those of Chapman & Cowling (1939, pp. 325–8).

by (7). Thus the velocity of diffusion is given by

$$\bar{c}_1 - \bar{c}_2 = -\frac{p\rho\tau}{\rho_1\rho_2}\frac{(\boldsymbol{d}_{12} - H^{-1}\omega\tau\boldsymbol{H}\wedge\boldsymbol{d}_{12})}{1+\omega^2\tau^2}.$$

This equation is identical with (4). That is, the discrepancy between the free-path and velocity-distribution results is removed when the mean peculiar velocity of molecules just after collision is assumed to be, not zero, but $\boldsymbol{c}'$, given by (11).

Because the diffusion velocity of molecules of one constituent gas of a mixture is, in part, transferred to molecules of other constituents by collisions, the diffusion velocity of any constituent does not depend wholly on factors affecting only that constituent. This deprives the free-path method of its chief simplicity. The free-path discussion cannot, moreover, itself determine the collision-interval τ. We therefore use only the velocity-distribution method in later work.

4. BINARY GAS; SECOND APPROXIMATION

Equation (6) gives only a first approximation to the velocity of diffusion in a binary gas. The first approximation, like other first approximations of the velocity-distribution theory, gives exact results for Maxwellian molecules (molecules repelling each other according to an inverse fifth power law); the free-path assumption of a collision-interval independent of the molecular velocity is essentially equivalent to assuming Maxwellian molecules. In the absence of a magnetic field the first approximation usually gives results correct to within 2 or 3 % for non-Maxwellian molecules; this, however, is not true when the force of interaction between molecules follows the inverse-square law. For this law, when m_1/m_2 is large, and interaction between pairs of molecules m_2 is neglected, the first approximation to the diffusion velocity is only about 0·3 of the exact value (Chapman & Cowling 1939, p. 196). Since in an ionized gas the forces between molecules follow the inverse-square law, I have considered a second approximation to the velocity of diffusion for this law.

The method of calculation is indicated in Chapman & Cowling (1939, pp. 333–4); here only the results are quoted. The velocity of diffusion is given, to any order of approximation, by an equation of the form

$$\bar{C}_1 - \bar{C}_2 = -\frac{p\rho\tau}{\rho_1\rho_2}(\alpha\boldsymbol{d}_{12} + H^{-1}\beta\boldsymbol{H}\wedge\boldsymbol{d}_{12}). \tag{14}$$

The first approximation (6) corresponds to $\alpha = \alpha_0$, $\beta = \beta_0$, where

$$\alpha_0 + i\beta_0 = 1/(1+i\omega\tau). \tag{15}$$

If $\omega\tau$ is large (magnetic field large, density small), the second (and all later) approximations give values of α, β indistinguishable from α_0, β_0. If $\omega\tau$ is small, the second approximation gives a value of α greater than α_0, and a value of β numerically greater than β_0 save when both of these two quantities are small. It gives α not very much greater than α_0 if m_1, m_2 are not very different;* e.g. if $m_1/m_2 = \frac{3}{2}$, then taking in turn $n_1/n_2 = 2$, 1 and $\frac{1}{2}$, it is found that $\alpha/\alpha_0 = 1{\cdot}27$, $1{\cdot}21$ and $1{\cdot}25$, and $\beta/\beta_0 = 1{\cdot}65$, $2{\cdot}31$ and $0{\cdot}41$, the irregular variation of β/β_0 being due to the fact that β and β_0 both vanish between $n_1/n_2 = 1$ and $n_1/n_2 = \frac{1}{2}$. When m_1/m_2 is very large, α/α_0 may be considerably greater than unity, though even in this case it is not so great as when the mutual collisions of the light molecules are neglected. For example, if $n_1 = n_2$, we find $\alpha = 1{\cdot}96\alpha_0$, $\beta = 4{\cdot}88\beta_0$; if $n_1 = \frac{1}{10}n_2$ ($e_1 = -10e_2$, corresponding to a highly ionized gas), $\alpha = 3\alpha_0$, $\beta = 13{\cdot}3\beta_0$.

The theory of electrostatic interaction is, in any case, only approximate, and in most applications the data are very uncertain. It is therefore usually adequate to use the first approximation in considering electric currents due to transport of ions, in which the molecular masses are comparable; this is especially true when large numbers of neutral molecules are present, since interaction with these does not demand such large corrections to α_0, β_0 as inverse-square law interaction. In considering the currents due to electrons, the error of the first approximation is too large to be ignored when $\omega\tau$ is small. The figures quoted above suggest that in this case, since $\beta_0 \propto \tau$, it is sufficient in (14) to use the first approximation with an increased value of τ.

5. MULTIPLE GAS

Now consider diffusion in a gas with several constituents. The mass-velocity $\mathbf{c}_0$ is supposed to be uniform, so as to remove the viscosity terms from the equations. If the velocity-distribution function f_r is regarded as a function of $\mathbf{C}_r$, $\mathbf{r}$ and t, the Boltzmann equation which it satisfies is of the form

$$\frac{Df_r}{Dt} + \mathbf{C}_r \cdot \frac{\partial f_r}{\partial \mathbf{r}} + \left(\mathbf{F}_r - \frac{D\mathbf{c}_0}{Dt}\right) \cdot \frac{\partial f_r}{\partial \mathbf{C}_r} + \frac{e_r}{m_r}(\mathbf{C}_r \wedge \mathbf{H}) \cdot \frac{\partial f_r}{\partial \mathbf{C}_r} = \sum_s \iint \{f_r' f_s' - f_r f_s\} k_{rs} \, d\mathbf{k} \, d\mathbf{c}_s$$

(cf. Chapman & Cowling 1939, p. 323, equation (3)). The first approximation to the solution of this is the Maxwellian function

$$f_r^{(0)} = n_r \left(\frac{m_r}{2\pi kT}\right)^{\frac{3}{2}} \exp\left[-\frac{m_r C_r^2}{2kT}\right].$$

The second approximation is $f_r^{(0)}(1+\phi_r)$, where

$$\frac{D_0 f_r^{(0)}}{Dt} + \mathbf{C}_r \cdot \frac{\partial f_r^{(0)}}{\partial \mathbf{r}} + \left(\mathbf{F}_r - \frac{D_0 \mathbf{c}_0}{Dt}\right) \cdot \frac{\partial f_r^{(0)}}{\partial \mathbf{C}_r} + \frac{e_r}{m_r} f_r^{(0)} (\mathbf{C}_r \wedge \mathbf{H}) \cdot \frac{\partial \phi_r}{\partial \mathbf{C}_r} = -\sum_s I_{rs}(\phi_r + \phi_s) \qquad (16)$$

and

$$I_{rs}(\phi_r + \phi_s) \equiv \iint f_r^{(0)} f_s^{(0)} (\phi_r + \phi_s - \phi_r' - \phi_s') \, k_{rs} \, d\mathbf{k} \, d\mathbf{c}_s. \qquad (17)$$

---

* The constant A appearing in the general expressions (Chapman & Cowling 1939, p. 164) is less, but not much less, than 0·4 for inverse-square law interaction; its value is here taken as 0·375.

In evaluating the first and third terms on the left of (16), the following first approximations are used to the equation of continuity for the rth gas, and to the equations of energy and momentum:

$$\frac{D_0 n_r}{Dt} = 0, \quad \frac{D_0 T}{Dt} = 0, \quad \rho\frac{D_0 \boldsymbol{c}_0}{Dt} = \sum_r \rho_r \boldsymbol{F}_r + \boldsymbol{j}\wedge\boldsymbol{H} - \frac{\partial p}{\partial \boldsymbol{r}}. \tag{18}$$

In the last equation, $\boldsymbol{j}$ is taken to be the current given by the second approximation to f_r. The first three terms on the left of (16) accordingly reduce to

$$\begin{aligned} f_r^{(0)}&\left\{\boldsymbol{C}_r.\frac{\partial \log f_r^{(0)}}{\partial \boldsymbol{r}} + \left(\boldsymbol{F}_r - \frac{D_0\boldsymbol{c}_0}{Dt}\right).\frac{\partial \log f_r^{(0)}}{\partial \boldsymbol{C}_r}\right\} \\ &= f_r^{(0)}\boldsymbol{C}_r.\left\{\frac{1}{n_r}\frac{\partial n_r}{\partial \boldsymbol{r}} - \frac{m_r}{kT}\left(\boldsymbol{F}_r - \frac{D_0\boldsymbol{c}_0}{Dt}\right)\right\} \\ &= f_r^{(0)}\boldsymbol{C}_r.\left\{\frac{n}{n_r}\boldsymbol{d}_r + \frac{m_r}{\rho kT}\boldsymbol{j}\wedge\boldsymbol{H}\right\}. \end{aligned}$$

In this, $\boldsymbol{d}_r$ denotes the vector

$$\boldsymbol{d}_r = \frac{\rho_r}{p}\left(\frac{D_0\boldsymbol{c}_0}{Dt} - \boldsymbol{F}_r - \frac{1}{\rho}\boldsymbol{j}\wedge\boldsymbol{H} + \frac{1}{\rho_r}\frac{\partial p_r}{\partial \boldsymbol{r}}\right) \tag{19}$$

$$= \frac{\partial}{\partial \boldsymbol{r}}\left(\frac{n_r}{n}\right) - \frac{(n\rho_r - n_r\rho)}{n\rho p}\frac{\partial p}{\partial \boldsymbol{r}} - \frac{1}{\rho p}\sum_s \rho_r\rho_s(\boldsymbol{F}_r - \boldsymbol{F}_s), \tag{19'}$$

so that

$$\sum_r \boldsymbol{d}_r = 0. \tag{20}$$

In general the force $m_r\boldsymbol{F}_r$ on a molecule is the sum of a gravitational force $m_r\boldsymbol{g}$ and an electric force $e_r\boldsymbol{E}$ (where $\boldsymbol{E}$ may include a part $\boldsymbol{c}_0\wedge\boldsymbol{H}$ due to mass motion in the magnetic field). Thus (19′) reduces to

$$\boldsymbol{d}_r = \frac{1}{p}\left(\frac{\partial p_r}{\partial \boldsymbol{r}} - \frac{\rho_r}{\rho}\frac{\partial p}{\partial \boldsymbol{r}} - n_r e_r \boldsymbol{E}\right). \tag{21}$$

Equation (16) now takes the form

$$\frac{n}{n_r}f_r^{(0)}\boldsymbol{C}_r.\boldsymbol{d}_r = -f_r^{(0)}\boldsymbol{C}_r.\left\{\frac{m_r}{\rho kT}\boldsymbol{j}\wedge\boldsymbol{H} + \frac{e_r}{m_r}\boldsymbol{H}\wedge\frac{\partial\phi_r}{\partial\boldsymbol{C}_r}\right\} - \sum_s I_{rs}(\phi_r + \phi_s). \tag{22}$$

The solution of this is of the form

$$\phi_r = -n\sum_s \boldsymbol{d}_s.\{D_{rs}^{\mathrm{I}}\boldsymbol{C}_r + D_{rs}^{\mathrm{II}}\boldsymbol{C}_r\wedge\boldsymbol{H}\}, \tag{23}$$

where D_{rs}^{I}, D_{rs}^{II} are functions of the scalars C_r and H. The corresponding velocity of diffusion, $\overline{\boldsymbol{C}}_r$, is given by

$$\begin{aligned} n_r\overline{\boldsymbol{C}}_r &= \int f_r^{(0)}\phi_r\boldsymbol{C}_r d\boldsymbol{c}_r \\ &= -\tfrac{1}{3}n\int f_r^{(0)}C_r^2\sum_s(D_{rs}^{\mathrm{I}}\boldsymbol{d}_s + D_{rs}^{\mathrm{II}}\boldsymbol{H}\wedge\boldsymbol{d}_s)\,d\boldsymbol{c}_r. \end{aligned} \tag{24}$$

Hence, in the absence of a magnetic field, each velocity of diffusion has components parallel to each vector $\boldsymbol{d}_s$. This implies that any one gas in the mixture may acquire a velocity of diffusion as a result of factors not directly affecting it, e.g. as a result of electric forces acting on other gases in the mixture, even when its own molecules are uncharged. As shown in § 3, this 'induced' diffusion is due to the transfer of the diffusion velocity of one gas to others at collisions. The introduction of a magnetic field alters the magnitude of the component of $\overline{\boldsymbol{C}}_r$ parallel to any $\boldsymbol{d}_s$, and also produces new components, perpendicular to $\boldsymbol{H}$ and to each $\boldsymbol{d}_s$.

Similar remarks apply to the current-density $\boldsymbol{j}$, given by

$$\begin{aligned}\boldsymbol{j} &= \sum_r n_r e_r \overline{\boldsymbol{C}}_r \\ &= -n \sum_s (L_s^{\mathrm{I}} \boldsymbol{d}_s + L_s^{\mathrm{II}} \boldsymbol{H} \wedge \boldsymbol{d}_s),\end{aligned} \tag{25}$$

where $$L_s^{\mathrm{I}} = \tfrac{1}{3} \sum_r \int f_r^{(0)} e_r C_r^2 D_{rs}^{\mathrm{I}} d\boldsymbol{c}_s, \quad L_s^{\mathrm{II}} = \tfrac{1}{3} \sum_r \int f_r^{(0)} e_r C_r^2 D_{rs}^{\mathrm{II}} d\boldsymbol{c}_s. \tag{26}$$

If the current is due solely to an electric field $\boldsymbol{E}$,

$$\boldsymbol{d}_r = -n_r e_r \boldsymbol{E}/p$$

and so $$\boldsymbol{j} = \sigma^{\mathrm{I}} \boldsymbol{E} + H^{-1} \sigma^{\mathrm{II}} \boldsymbol{H} \wedge \boldsymbol{E}, \tag{27}$$

where $$\sigma^{\mathrm{I}} = \frac{1}{kT} \Sigma n_s e_s L_s^{\mathrm{I}}, \quad \sigma^{\mathrm{II}} = \frac{H}{kT} \Sigma n_s e_s L_s^{\mathrm{II}}. \tag{28}$$

Equation (27) implies that, in the presence of a transverse magnetic field, an electric field $\boldsymbol{E}$ produces currents of magnitude $\sigma^{\mathrm{I}} E$ and $\sigma^{\mathrm{II}} E$, respectively parallel to $\boldsymbol{E}$ and perpendicular to both $\boldsymbol{H}$ and $\boldsymbol{E}$. σ^{I} and σ^{II} may be called the direct and transverse conductivities.

6. FIRST APPROXIMATION TO DIFFUSION VELOCITIES

Substituting for ϕ_r and $\boldsymbol{j}$ from (23) and (25) into (22), then

$$\begin{aligned}\frac{1}{n_r} f_r^{(0)} \boldsymbol{C}_r . \boldsymbol{d}_r = f_r^{(0)} \frac{m_r \boldsymbol{C}_r}{\rho kT} . \sum_s (-L_s^{\mathrm{I}} \boldsymbol{H} \wedge \boldsymbol{d}_s + H^2 L_s^{\mathrm{II}} \boldsymbol{d}_s) + \frac{e_r}{m_r} f_r^{(0)} \boldsymbol{C}_r . \sum_s (D_{rs}^{\mathrm{I}} \boldsymbol{H} \wedge \boldsymbol{d}_s - H^2 D_{rs}^{\mathrm{II}} \boldsymbol{d}_s) \\ + \sum_q I_{rq} \left(\sum_s \boldsymbol{d}_s . \{ D_{rs}^{\mathrm{I}} \boldsymbol{C}_r + D_{rs}^{\mathrm{II}} \boldsymbol{C}_r \wedge \boldsymbol{H} + D_{qs}^{\mathrm{I}} \boldsymbol{C}_q + D_{qs}^{\mathrm{II}} \boldsymbol{C}_q \wedge \boldsymbol{H} \} \right).\end{aligned} \tag{29}$$

In this, the vectors $\boldsymbol{d}_s$ are not independent, because of (20). Equally, however, the functions D_{rs}^{I}, D_{rs}^{II} are not uniquely defined so far; for any given r, the value of ϕ_r given by (23) is unaffected if each D_{rs}^{I} or D_{rs}^{II} is increased by the same amount. To make D_{rs}^{I} and D_{rs}^{II} definite, suppose that for a special value s' of s

$$D_{rs'}^{\mathrm{I}} = 0 = D_{rs'}^{\mathrm{II}}.$$

Suppose also that (20) is used to express $\boldsymbol{d}_{s'}$ in terms of the other vectors $\boldsymbol{d}_s$. These other vectors are independent; thus one can afterwards equate coefficients of $\boldsymbol{d}_s$ and $\boldsymbol{H}\wedge\boldsymbol{d}_s$ on opposite sides of (29). This gives

$$\left.\begin{aligned}\frac{1}{n_r}f_r^{(0)}\boldsymbol{C}_r(\delta_{rs}-\delta_{rs'}) &= H^2 f_r^{(0)}\boldsymbol{C}_r\left(\frac{m_r}{\rho kT}L_s^{\text{II}}-\frac{e_r}{m_r}D_{rs}^{\text{II}}\right)+\sum_q I_{rq}(D_{rs}^{\text{I}}\boldsymbol{C}_r+D_{qs}^{\text{I}}\boldsymbol{C}_q),\\ 0 &= -f_r^{(0)}\boldsymbol{C}_r\left(\frac{m_r}{\rho kT}L_s^{\text{I}}-\frac{e_r}{m_r}D_{rs}^{\text{I}}\right)+\sum_q I_{rq}(D_{rs}^{\text{II}}\boldsymbol{C}_r+D_{qs}^{\text{II}}\boldsymbol{C}_q).\end{aligned}\right\} \tag{30}$$

Now put $$\varDelta_{rs} = D_{rs}^{\text{I}}+iHD_{rs}^{\text{II}}, \quad \varLambda_s = L_s^{\text{I}}+iHL_s^{\text{II}}. \tag{31}$$

Then equations (30) can be replaced by the complex equation

$$\frac{1}{n_r}f_r^{(0)}\boldsymbol{C}_r(\delta_{rs}-\delta_{rs'}) = iHf_r^{(0)}\boldsymbol{C}_r\left(-\frac{m_r}{\rho kT}\varLambda_s+\frac{e_r}{m_r}\varDelta_{rs}\right)+\sum_q I_{rq}(\varDelta_{rs}\boldsymbol{C}_r+\varDelta_{qs}\boldsymbol{C}_q). \tag{32}$$

In this, q and r can take all values; s takes all values other than s'.

Only a first approximation to the solution of (32) will be considered, similar to that yielding (6) for a binary gas. This is found by regarding $\varDelta_{rs}$ as a quantity independent of $\boldsymbol{C}_r$, to be determined, not from (32), but from the equations obtained by taking the scalar product of $m_r\boldsymbol{C}_r$ with (32), and integrating over all values of $\boldsymbol{c}_r$. In the notation of Chapman & Cowling (1939, p. 86, equation (9)), the resulting equations are

$$3kT(\delta_{rs}-\delta_{rs'}) = 3iH\left(-\frac{\rho_r}{\rho}\varLambda_s+\frac{e_r}{m_r}n_r kT\varDelta_{rs}\right)+\sum_q n_q n_r[m_r\boldsymbol{C}_r, \varDelta_{rs}\boldsymbol{C}_r+\varDelta_{qs}\boldsymbol{C}_q]_{qr}. \tag{33}$$

Also from (24), taking $\varDelta_{rs}$ (and therefore D_{rs}^{I}, D_{rs}^{II}) as independent of $\boldsymbol{C}_r$,

$$n_r\overline{\boldsymbol{C}}_r = -n\frac{p_r}{m_r}\sum_s(D_{rs}^{\text{I}}\boldsymbol{d}_s+D_{rs}^{\text{II}}\boldsymbol{H}\wedge\boldsymbol{d}_s). \tag{34}$$

Remembering that $n\boldsymbol{d}_s = \partial n_s/\partial\boldsymbol{r}$ for an inhomogeneous gas at rest under no forces, it is seen that $p_r D_{rs}^{\text{I}}/m_r$, $Hp_r D_{rs}^{\text{II}}/m_r$ may be regarded as generalized diffusion coefficients for direct and transverse diffusion.

The collision integral appearing as the last term on the right of (33) can be evaluated, using the formulae of Chapman & Cowling (1939, § 9·8). The result is

$$\begin{aligned}[m_r\boldsymbol{C}_r, \varDelta_{rs}\boldsymbol{C}_r+\varDelta_{qs}\boldsymbol{C}_q]_{qr} &= \frac{16kT}{m_q+m_r}(m_q\varDelta_{rs}-m_r\varDelta_{qs})\varOmega_{qr}^{(1)}(1)\\ &= \frac{3k^2T^2}{(n_q+n_r)D_{qr}}\left\{\frac{\varDelta_{rs}}{m_r}-\frac{\varDelta_{qs}}{m_q}\right\},\end{aligned}$$

where D_{qr} would be the ordinary coefficient of mutual diffusion of the qth and rth gases in the absence of the other gases. This result can be expressed in terms of an effective collision-interval τ_{qr} for molecules of the qth and rth gases (similar to the τ of (8)), such that

$$\tau_{qr} = \frac{m_q m_r(n_q+n_r)}{(\rho_q+\rho_r)kT}D_{qr}. \tag{35}$$

It is, however, more convenient to use θ_{qr}, defined by

$$\theta_{qr} = \frac{\rho_q \rho_r}{(\rho_q + \rho_r)\,\tau_{qr}} = \frac{n_q n_r kT}{(n_q + n_r)\,D_{qr}}. \tag{36}$$

Then
$$[m_r C_r, \varDelta_{rs} C_r + \varDelta_{qs} C_q]_{qr} = \frac{3kT}{n_q n_r}\theta_{qr}\left(\frac{\varDelta_{rs}}{m_r} - \frac{\varDelta_{qs}}{m_q}\right)$$

and equation (33) becomes

$$\delta_{rs} - \delta_{rs'} = iH\left(-\frac{\rho_r}{\rho kT}\Lambda_s + e_r n_r \frac{\varDelta_{rs}}{m_r}\right) + \sum_q \theta_{qr}\left(\frac{\varDelta_{rs}}{m_r} - \frac{\varDelta_{qs}}{m_q}\right). \tag{37}$$

If this relation is summed over all values of r, it is found that

$$\Lambda_s = kT \sum_r e_r n_r \varDelta_{rs}/m_r. \tag{38}$$

This equation could also be obtained directly from (26) and (31).

In addition to (37), further equations are satisfied by the quantities $\varDelta_{rs}$, in virtue of the momentum equation (10). Substituting for $\overline{C}_r$ from (34) into this equation,

$$0 = \sum_r \rho_r \varDelta_{rs}/m_r. \tag{39}$$

Equations (37) and (39) suffice to determine the quantities $\varDelta_{rs}$ and Λ_s. The determinantal solution is of the form

$$H\Lambda_s/(\rho kT) = i\{\mathscr{L}_{N+1,s} - \mathscr{L}_{N+1,s'}\}/\mathscr{L}, \tag{40}$$

$$\varDelta_{rs}/m_r = \{\mathscr{L}_{r,s} - \mathscr{L}_{r,s'}\}/\mathscr{L}. \tag{41}$$

Here $\mathscr{L}$ is a symmetric determinant with $N+1$ rows and columns ($N =$ number of gases), the elements l_{rs} of which are given by

$$\left.\begin{aligned} l_{rs} &= -\theta_{rs} + \delta_{rs}\Big(\sum_q \theta_{qr} + e_r n_r iH\Big) \quad (r, s \leqslant N),\\ l_{N+1,r} &= l_{r,N+1} = \rho_r \quad (r \leqslant N), \qquad l_{N+1,N+1} = 0, \end{aligned}\right\} \tag{42}$$

and $\mathscr{L}_{rs}$ is the cofactor of l_{rs} in $\mathscr{L}$.

Equation (41) gives the formal solution to the problem of diffusion, since the generalized coefficients of diffusion $p_r D^{\mathrm{I}}_{rs}/m_r$, $Hp_r D^{\mathrm{II}}_{rs}/m_r$ are the real and imaginary parts of $p_r \varDelta_{rs}/m_r$. Similarly, (40) formally gives the electric currents (cf. (25), (31)). The direct and transverse conductivities are given by

$$\begin{aligned} \sigma^{\mathrm{I}} + i\sigma^{\mathrm{II}} &= (kT)^{-1} \sum_s n_s e_s \Lambda_s \\ &= \frac{i\rho}{H}\frac{\mathscr{L}_e}{\mathscr{L}}, \end{aligned} \tag{43}$$

where $\mathscr{L}_e$ is got from $\mathscr{L}$ by replacing the terms ρ_r of the last row by $n_r e_r$.

7. TERNARY MIXTURE

For a mixture of three gases, the determinant $\mathscr{L}$ becomes

$$\mathscr{L} = \begin{vmatrix} e_1 n_1 iH + \theta_{12} + \theta_{13}, & -\theta_{12}, & -\theta_{13}, & \rho_1 \\ -\theta_{12}, & e_2 n_2 iH + \theta_{23} + \theta_{12}, & -\theta_{23}, & \rho_2 \\ -\theta_{13}, & -\theta_{23}, & e_3 n_3 iH + \theta_{13} + \theta_{23}, & \rho_3 \\ \rho_1, & \rho_2, & \rho_3, & 0 \end{vmatrix}$$

$$= H^2 \Sigma \rho_1^2 e_2 n_2 e_3 n_3 - iH\rho \Sigma \theta_{23} n_1 e_1 (\rho - 2\rho_1) - \rho^2 (\theta_{12}\theta_{13} + \theta_{23}\theta_{12} + \theta_{13}\theta_{23}) \tag{44}$$

using (2). The cofactors of the terms of $\mathscr{L}$ can readily be found. The determinant $\mathscr{L}_e$ is given by

$$\mathscr{L}_e = H^2 \rho n_1 n_2 n_3 e_1 e_2 e_3 + iH\rho \Sigma n_1^2 e_1^2 \theta_{23}. \tag{45}$$

The general expressions for the electric currents and the conductivity can be simplified in certain special cases, which will now be discussed in detail.

Case I. Slightly ionized gas. Suppose that the first gas supplies nearly all of the molecules and of the total mass of the mixture. Since the gas as a whole is supposed to be electrically neutral, this implies in practice that $e_1 = 0$, $n_2 e_2 = -n_3 e_3$. Suppose, moreover, that nearly all of the collisions of molecules m_2, m_3 are with molecules m_1 (it is necessary to regard this as a separate assumption, because the effective collision-radius for the collisions of pairs of charged molecules is very large). Analytically, the second assumption implies that θ_{23} is small compared with θ_{12} and θ_{13}. Neglecting θ_{23}/θ_{12}, θ_{23}/θ_{13}, ρ_2/ρ_1, ρ_3/ρ_1,

$$\mathscr{L} = H^2 \rho^2 n_2 e_2 n_3 e_3 - iH\rho^2 (e_2 n_2 \theta_{13} + e_3 n_3 \theta_{12}) - \rho^2 \theta_{12} \theta_{13},$$

$$\mathscr{L}_e = -iH\rho n_2 e_2 n_3 e_3 (\theta_{12} + \theta_{13}).$$

Hence

$$\sigma^{\mathrm{I}} + i\sigma^{\mathrm{II}} = \frac{n_2^2 e_2^2}{\theta_{12} + iHe_2 n_2} + \frac{n_3^2 e_3^2}{\theta_{13} + iHe_3 n_3},$$

or, since (36) now gives $\theta_{12} = \rho_2/\tau_{12}$, $\theta_{13} = \rho_3/\tau_{13}$,

$$\sigma^{\mathrm{I}} + i\sigma^{\mathrm{II}} = \frac{n_2 e_2^2 \tau_{12}}{m_2(1 + i\omega_2 \tau_{12})} + \frac{n_3 e_3^2 \tau_{13}}{m_3(1 + i\omega_3 \tau_{13})}. \tag{46}$$

Thus the conductivity can be divided into two independent parts, one arising from the second gas, one from the third; (46) is, in fact, identical with the result obtained from the (uncorrected) free path-formula (4). More generally, in any mixture of N constituents, nearly all of whose mass and molecules are supplied by one gas, the correct conductivity can be deduced from (4). This is to be expected on general grounds. The uncorrected free-path theory is inadequate for a general mixture because it neglects the transfer of diffusion-velocity from one gas to another at collisions. But here the preponderant gas provides most of the mass, and has a negligible velocity of diffusion relative to the mass; it neither imparts nor receives any appreciable velocity of diffusion at collisions.

When diffusion is due to factors other than simply an electric field, the electric current is given by

$$\boldsymbol{j} = -\frac{e_2 p\tau_{12}}{m_2}\frac{(\boldsymbol{d}_2 - H^{-1}\omega_2\tau_{12}\boldsymbol{H}\wedge\boldsymbol{d}_2)}{1+\omega_2^2\tau_{12}^2} - \frac{e_3 p\tau_{13}}{m_3}\frac{(\boldsymbol{d}_3 - H^{-1}\omega_3\tau_{13}\boldsymbol{H}\wedge\boldsymbol{d}_3)}{1+\omega_3^2\tau_{13}^2}, \tag{47}$$

where $\boldsymbol{d}_2$, $\boldsymbol{d}_3$ are given by (21). This equation also is very similar to free-path results.

Case II. Electrons present. Consider a mixture in which the molecules of the first two gases are different kinds of ions, while those of the third are electrons, the number-densities of the three gases being similar in order of magnitude. Only the lowest powers of the small quantities m_3/m_1, m_3/m_2 in our expressions are retained. In (36),

$$(n_q + n_r)\,D_{qr} \propto \left(\frac{m_q+m_r}{m_q m_r}\right)^{\frac{1}{2}}, \quad \theta_{qr} \propto n_q n_r \left(\frac{m_q m_r}{m_q+m_r}\right)^{\frac{1}{2}}.$$

Hence θ_{12} is large compared with θ_{13} or θ_{23}, but small compared with $\rho_2\theta_{13}/\rho_3$ or $\rho_1\theta_{23}/\rho_3$. Thus approximately

$$\begin{aligned}\mathscr{L} &= H^2 e_3 n_1 n_2 n_3 (e_2\rho_1 m_1 + e_1\rho_2 m_2) - iH\rho^2 n_3 e_3\theta_{12} - \rho^2\theta_{12}(\theta_{13}+\theta_{23})\\ &\doteqdot -(\theta_{13}+\theta_{23}+iHe_3n_3)\,(\rho^2\theta_{12} + iHn_1n_2(e_2\rho_1m_1 + e_1\rho_2m_2))\end{aligned}$$

$$\begin{aligned}\mathscr{L}_e &= H^2\rho n_1 e_1 n_2 e_2 n_3 e_3 + iH\rho n_3^2 e_3^2\theta_{12}\\ &\doteqdot iH\rho^{-1}\{n_3^2e_3^2(\rho^2\theta_{12} + iHn_1n_2(e_2\rho_1m_1 + e_1\rho_2n_2))\\ &\qquad + n_1^2n_2^2(e_1m_2 - e_2m_1)^2\,(\theta_{13}+\theta_{23}+iHe_3n_3)\},\end{aligned}$$

and so
$$\sigma^{\mathrm{I}} + i\sigma^{\mathrm{II}} = \frac{n_3^2e_3^2}{\theta_{13}+\theta_{23}+iHn_3e_3} + \frac{n_1^2n_2^2(e_1m_2-e_2m_1)^2}{\rho^2\theta_{12}+iHn_1n_2(e_2\rho_1m_1+e_1\rho_2m_2)}. \tag{48}$$

Thus the conductivity can again be divided into two parts, one due to the diffusion of electrons relative to ions, and the other to the motion of one set of ions relative to the other. The electron conductivity does not depend on the mutual collisions of ions; since here approximately $\theta_{13} = \rho_3/\tau_{13}$, $\theta_{23} = \rho_3/\tau_{23}$, it is given by

$$\sigma_e^{\mathrm{I}} + i\sigma_e^{\mathrm{II}} = \frac{n_3e_3^2}{m_3(\tau_{13}^{-1}+\tau_{23}^{-1}) + iHe_3},$$

which is precisely the expression given by free-path theory. The ion conductivity is unaffected by electron collisions; if $\rho\omega_{12} = \rho_1\omega_2 + \rho_2\omega_1$, it is given by

$$\sigma_i^{\mathrm{I}} + i\sigma_i^{\mathrm{II}} = \frac{n_1n_2(e_1m_2-e_2m_1)^2\tau_{12}}{\rho m_1m_2(1+i\omega_{12}\tau_{12})},$$

which is precisely the conductivity found for a binary gas from (6).

Similarly, when other factors beside the electric field produce diffusion, the total electric current is the sum of independent ion and electron contributions. In fact

$$\begin{aligned}\boldsymbol{j} = &-\frac{e_3p}{m_3}\frac{((\tau_{13}^{-1}+\tau_{23}^{-1})\,\boldsymbol{d}_3 - H^{-1}\omega_3\boldsymbol{H}\wedge\boldsymbol{d}_3)}{((\tau_{13}^{-1}+\tau_{23}^{-1})^2+\omega_3^2)}\\ &+\frac{(m_1e_2-m_2e_1)\,p\tau_{12}(\rho_2\boldsymbol{d}_1 - \rho_1\boldsymbol{d}_2 - H^{-1}\omega_{12}\tau_{12}\boldsymbol{H}\wedge(\rho_2\boldsymbol{d}_1-\rho_1\boldsymbol{d}_2))}{m_1m_2\rho(1+\omega_{12}^2\tau_{12}^2)},\end{aligned} \tag{49}$$

where in the present case, by (21),

$$\boldsymbol{d}_3 = \frac{1}{p}\left(\frac{\partial p_3}{\partial r} - n_3 e_3 \boldsymbol{E}\right), \tag{50}$$

$$\rho_2 \boldsymbol{d}_1 - \rho_1 \boldsymbol{d}_2 = \frac{\rho_1 \rho_2}{p}\left(\frac{1}{\rho_1}\frac{\partial p_1}{\partial r} - \frac{e_1}{m_1}\boldsymbol{E} - \frac{1}{\rho_2}\frac{\partial p_2}{\partial r} + \frac{e_2}{m_2}\boldsymbol{E}\right). \tag{51}$$

Equations (48) and (49) cannot always be applied to a mixture in which one gas —say the first—has uncharged molecules. The effective collision-radii are much greater for pairs of ions than for ions with uncharged molecules, and so in this case it is permissible to nelgect θ_{23}/θ_{12} only when the uncharged molecules are numerous.

Generally, for a mixture of N gases, when electrons and ions of both signs are present, the conductivity and current can be divided into independent electron and ion contributions. The electron contribution can be calculated from free-path formulae; the collision-interval to be used in these formulae is that given by the first approximation of § 6, somewhat increased where necessary, in the way indicated in § 4. The accuracy of the first approximation is, however, usually adequate when a large number of neutral molecules are present; not only is the effect of electron collisions with ions 'diluted' by the presence of the neutral molecules, but the correction to the effective electron-ion collision-interval is reduced by the collisions with the neutral molecules. Ion conductivity cannot in general be expressed by free-path formulae (uncorrected) for the reasons noted in § 3. When a large number of neutral molecules is present, it can be so expressed; but when an adequate number of ions is present, the mutual collision of pairs of ions becomes important, and reduces the total conductivity below the sum of the (free-path) contributions of the different sets of charged particles present.

8. SOLAR APPLICATIONS

Sunspots possess magnetic fields of order 2000 gauss; the sun is also stated to possess a general magnetic field of order 25 gauss. The spot field produces marked changes in the conductivity in the surface layers.

In the sun, the contribution of ions to the conductivity is always smaller than that of electrons, and is comparable with it only in extreme cases, namely, when the pressure is low and the magnetic field large. Since in any case the conductivity then is small, the ion conductivity may in practice be ignored.

To study the electron conductivity, it is convenient to introduce a mean collision distance for molecules of the rth and sth gases. This is defined as the sum of the radii of elastic spherical molecules of the same masses, giving the same collision-function θ_{rs} and collision-interval τ_{rs}. The inverse-square law of interaction of electrons with ions gives an expression for θ_{rs} depending on a slowly varying function $A_1(2)$ of the density and temperature (Chapman & Cowling 1939, p. 179). In the sun's atmosphere $A_1(2)$ lies between 10 and 11; we here use the mean value $A_1(2) = 10{\cdot}6$. Then the mean collision-distance for electrons with Z-fold ionized atoms is

$$4{\cdot}7 \times 10^{-7} Z(5750/T)\ \text{cm}.$$

As mobility measurements indicate, the collision-distance for electrons with neutral molecules is probably not greater than 2×10^{-8} cm. Hence, unless neutral atoms are, say, 100 times as numerous as ions in the sun's atmosphere, it is permissible to neglect collisions with neutral atoms. In fact, neutral atoms (hydrogen and helium) are usually supposed to be only about 30 times as numerous as ionized atoms (metals) in the surface layers of the sun's atmosphere; they disappear completely at a depth of a few hundred km. below the visible layers.

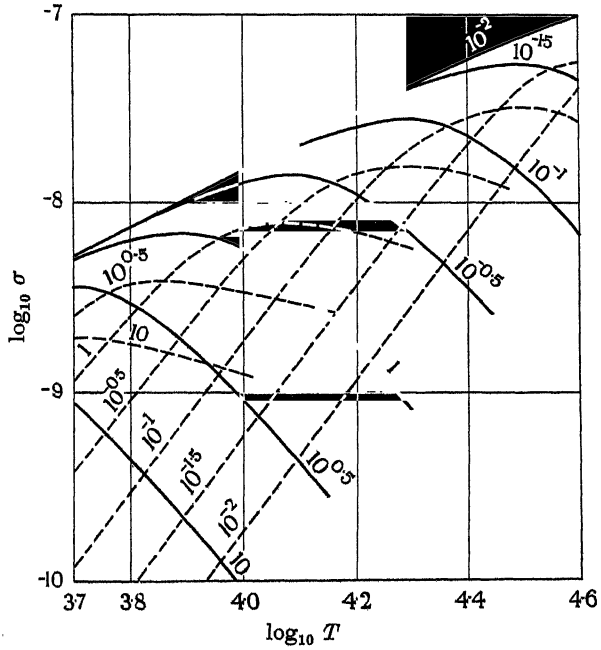


FIGURE 1. Graphs of $\log_{10} \sigma^{\mathrm{I}}$ (continuous curves) and $\log_{10} \sigma^{\mathrm{II}}$ (broken curves) against $\log_{10} T$ for certain values of H/p_e shown against the corresponding curves.

Neglecting collisions with neutral atoms, a general formula for the (electron) conductivity may be given. Suppose several types of ion are present; let the number-density of the rth type be n_r, and let it be Z_r times ionized. The mean degree of ionization $\bar{Z}$ is defined by

$$\bar{Z} = \Sigma n_r Z_r^2 / \Sigma n_r Z_r. \tag{52}$$

In terms of this

$$\sigma^{\mathrm{I}} + i\sigma^{\mathrm{II}} = \{6{\cdot}8 \times 10^{13} \bar{Z} T^{-\frac{3}{2}} - i 8{\cdot}6 \times 10^{3} HT/p_e\}^{-1} \text{ e.m.u.} \tag{53}$$

where p_e is the electron-pressure in dynes/cm.². The values of σ^{I} and σ^{II} calculated from (53) for surface layers ($\bar{Z} = 1$) are shown graphically in figure 1.

According to (53), σ^{II} is smaller than σ^{I}, and σ^{I} is not very greatly reduced by the magnetic field, if

$$H < \frac{6{\cdot}8 \times 10^{13}}{8{\cdot}6 \times 10^{3}} \frac{\bar{Z} p_e}{T^{\frac{5}{2}}} = 7{\cdot}9 \times 10^{9} \frac{\bar{Z} p_e}{T^{\frac{5}{2}}} \text{gauss.}$$

For example, if $\bar{Z} = 1, p_e = 40$ dynes/cm.², $T = 5750°$ K (taken as typical of the surface layers), this gives $H < 122$ gauss; if $\bar{Z} = 1, p_e = 2 \times 10^4$ dynes/cm.², $T = 17,000°$ K (taken as typical of the layer in which hydrogen ionization is nearly complete, say 400 km. lower), it gives $H < 4200$ gauss. Hence, if the magnetic field in the interior is never much greater than at the surface, its effect on σ^{I} is likely not to be large at a depth of more than 1000 km. below the sun's surface. Even at the surface, the effect on σ^{I} is not large save near a sunspot; but near the centre of a spot σ^{I} may be only a fraction of 1 % of its value in the absence of the magnetic field.

At great depths, and in general when σ^{II} is small compared with σ^{I}, approximately

$$\sigma^{\mathrm{I}} = 1{\cdot}47 \times 10^{-14} T^{\frac{3}{2}}/\bar{Z} \text{ e.m.u.} \tag{54}$$

The values given by (53) and (54) are first approximations. When σ^{II} is larger than σ^{I}, the first approximation involves little error. When $\sigma^{\mathrm{II}}/\sigma^{\mathrm{I}}$ is small, the first approximation can be roughly corrected by replacing $\bar{Z}$ by $\bar{Z}'$, where the following values of $\bar{Z}'$ are inferred from the second approximation quoted in § 4:

$\bar{Z}$	1	1·5	2	2·5	3	4	5	6	8	10	15
$\bar{Z}'$	0·51	0·67	0·83	0·98	1·14	1·44	1·73	2·02	2·60	3·18	4·65

Thus in surface layers, away from a spot, σ^{I} is of order 10^{-8} e.m.u.; at a level 500 km. lower, where hydrogen ionization is just complete, σ^{I} is of order 10^{-7} e.m.u.; at the centre of the sun it is of order 6×10^{-4} e.m.u.; and half-way along the radius, of order 4×10^{-5} e.m.u. (taking $\bar{Z}' = 3$ in the last two estimates).

In addition to conduction currents due to an applied electric force, currents arise in general whenever diffusion occurs. Electrons normally contribute the major part of these, as of conduction currents. Their magnitude can be found from the conductivity results by regarding the electrons as subject to an equivalent electric force $\boldsymbol{E}'$, where

$$\boldsymbol{E}' = \boldsymbol{E} + \frac{kT}{ep_e} \frac{\partial p_e}{\partial \boldsymbol{r}},$$

$-e$ being the electronic charge ($-1{\cdot}59 \times 10^{-20}$ e.m.u.).

9. MECHANICAL EFFECTS

The conductivity is, however, not the most important factor determining the currents which flow in the sun's outer layers. Because of the currents induced in it, a mass of gas, moving with velocity $\boldsymbol{c}_0$ across the lines of force, is subject to a retarding force $-\sigma^{\mathrm{I}} H^2 \boldsymbol{c}_0$ per unit volume. This destroys the momentum $\rho \boldsymbol{c}_0$ in a time of order $\rho(\sigma^{\mathrm{I}} H^2)^{-1}$. In the first thousand km. below the sun's surface, $\rho/\sigma^{\mathrm{I}} < 10$ g.cm.$^{-3}$/e.m.u.; hence, if $H > 10$ gauss, the time of decay of the motion is a fraction of a second. The motion of the gas can persist only if (*a*) it is maintained by an accelerating force (including, as a special case, a pressure gradient), or (*b*) the gas is polarized by its motion in the magnetic field, and an electrostatic field is set up, which neutralizes the field $\boldsymbol{c}_0 \wedge \boldsymbol{H}$; or (*c*) the interior of a moving mass of gas is shielded from the external magnetic field by currents induced in its surface layers.

In case (*a*), the motion maintained is such that the force $\boldsymbol{j} \wedge \boldsymbol{H}$ due to the induced current just balances the accelerating force. Thus the current is determined by the latter force; the conductivity merely determines the transverse velocity required to produce the current. This velocity is normally not large, as solar velocities go. For example, when $\sigma^{\text{II}}/\sigma^{\text{I}}$ can be neglected, gravity g, acting across a horizontal magnetic field, produces a vertical velocity $\rho g/H^2\sigma^{\text{I}}$; taking $\rho/\sigma^{\text{I}} < 10$ g.cm.$^{-3}$/e.m.u., $H > 10$ gauss, this is found to be less than 30 m./sec., as against velocities of order 1 km./sec. often met near a sunspot. Hence, unless polarization or electromagnetic shielding occur (cases (*b*) and (*c*)), motions in the outer layers of the sun are directed almost exactly along the lines of magnetic force.

In cases (*b*) and (*c*), the full electric force does not act on the gas. Electric polarization (case (*b*)) is able only partially to neutralize the field $\boldsymbol{c}_0 \wedge \boldsymbol{H}$ in the sun's atmosphere; the space-charge produced by polarization is dissipated through the surrounding material at a rate comparable with that at which it is produced. Thus polarization alone cannot greatly reduce the rate of decay of a motion across the lines of force. In the motion of a stream of ionized gas outward from the sun (assumed in certain theories of magnetic storms) the space-charge is dissipated less readily, because the stream is not surrounded by conducting material.

Electromagnetic shielding (case (*c*)) is important for a large mass of moving gas. If a mass of gas, in the interior of which there is no magnetic field, moves in a region where the field is H, the currents induced in its outer layers imply a retarding pressure, comparable with the electromagnetic stress $H^2/8\pi$, over its boundary. In unit time this destroys momentum equal to that of the gas within a distance $H^2(8\pi\rho c_0)^{-1}$ of the boundary. The distance in question is 400 km. if $H = 1000$ gauss, $\rho \doteq 10^{-8}$ g./cm.^{3}, $c_0 = 1$ km./sec., and so in this case motion across the lines of force would persist for only a few seconds even if the dimensions of the moving mass are of order 1000 km. That is, electromagnetic shielding is relatively unimportant in the surface layers near a sunspot. At great depths, or in regions where the magnetic field is small, however, shielding can be effective for relatively small moving volumes. It is, of course, not necessary to suppose that the magnetic field vanishes in the transported gas; this may carry its own magnetic field with it in its interior. Thus, for example, the field of a sunspot might be carried to the surface from a deep level by the convection of the sunspot material.

To summarize this discussion, motions across the lines of force can persist below the sun's surface because of electromagnetic shielding, and isolated clouds well above the surface may be protected from magnetic retardation by polarization. But in the surface layers, especially near sunspots, motion must be nearly along the lines of magnetic force. This undoubtedly is why prominences often end their existence by being sucked into a spot; why, in eruptions near spots, the ejected (largely ionized) gas is often seen, on slow-motion cinema records, to travel up and return down the same curved path; and why, in such eruptions, the gas shows little tendency to expand perpendicular to its direction of motion, in spite of thermal velocities of the atoms of order 10 km./sec. It is probable that the persistent forms of quiescent prominences are to be explained similarly.

10. APPLICATION TO THE EARTH'S UPPER ATMOSPHERE

The conductivity of the extreme upper atmosphere (above a height of 80 km., say) is important in considering the solar and lunar variations of the earth's magnetic field. To discuss the conductivity adequately, it is necessary to consider a mixture of four gases—neutral molecules, positive ions, negative ions, and electrons. These are distinguished respectively by the suffixes a, $+$, $-$, and e. The molecular weights of neutral molecules and ions are both taken to be 26. If oxygen is largely molecular, this choice of molecular weights results in an overestimate of the ion conductivity (the coefficient of mutual diffusion of molecules of masses m_r, m_s is proportional to $(m_r+m_s)^{\frac{1}{2}} m_r^{-\frac{1}{2}} m_s^{-\frac{1}{2}}$); if oxygen is atomic, there is an underestimate; since nitrogen is probably present in large quantities, the error should not, however, exceed 10 %.

The quantity $A_1(2)$ involved in electrostatic interaction is rather larger than in solar applications, because of the much smaller number-density of charged particles. Taking $A_1(2) = 17{\cdot}5$, it is found that

$$\theta_{+e}/n_+ = \theta_{-e}/n_- = 2{\cdot}84\times 10^{-26} n_e T^{-\frac{3}{2}}, \quad \theta_{+-} = 4{\cdot}37\times 10^{-24} n_+ n_- T^{-\frac{3}{2}}. \tag{55}$$

This corresponds to a mean collision-distance $1{\cdot}16\times 10^{-5}(300/T)$ cm. for pairs of charged particles.

For neutral molecules with ions of either sign, the collision-distance is assumed to be $5{\cdot}9\times 10^{-8}(300/T)^{\frac{1}{2}}$ cm. (suggested by the observed mobility of N_2^+ ions in N_2; Tyndall 1938). For electrons with neutral molecules, the collision-distance is taken to be 2×10^{-8} cm., as suggested by the observed mobility of electrons in air (Townsend & Tizard 1913; this work, however, chiefly referred to electrons with velocities much greater than thermal). Then

$$\theta_{a+}/n_+ = \theta_{a-}/n_- = 2{\cdot}25\times 10^{-32} n_a, \quad \theta_{ae} = 9{\cdot}35\times 10^{-37} n_a n_e T^{\frac{1}{2}}. \tag{56}$$

The maximum number-densities of electrons in the E, F_1 and F_2 layers, at noon at the equator near sunspot maximum, are about 2×10^5, $3{\cdot}5\times 10^5$, and $1{\cdot}8\times 10^6$/cm.^{3}. The ratios n_-/n_e, n_+/n_e are denoted by λ, $1+\lambda$; it is usually supposed that λ is about 100 in the E-layer, much smaller (though probably greater than unity) in F_1, and less than unity in F_2. Massey & Bates (1943) have recently suggested that λ is less than unity, even in the E-layer; but they consider only atomic oxygen, and molecular oxygen is likely to be important in the E-layer.

The number of neutral atoms is normally estimated from the effective collision-interval τ of electrons, as inferred from radio reflexion coefficients. Observations in England suggest a collision-frequency τ^{-1} about 6×10^4/sec. at the centre of the E-layer, 4×10^3/sec. for F_1, and 2×10^3/sec. for F_2; the corresponding number-densities of electrons are, say, 10^5, $2{\cdot}5\times10^5$ and 5×10^5/cm.^{3}. The collision-interval so found is precisely that effective in diffusion; it is given by

$$\begin{aligned}\tau^{-1} &= (\theta_{ae}+\theta_{+e}+\theta_{-e})/m_e n_e \\ &= 1{\cdot}8\times10^{-8}n_a\left(\frac{T}{300}\right)^{\frac{1}{2}}+6{\cdot}06\times10^{-3}(n_++n_-)\left(\frac{300}{T}\right)^{\frac{3}{2}}. \qquad (57)\end{aligned}$$

In spite of the relative fewness of ions, these must not be neglected in considering collision-intervals; because of the large collision-distance for ions with electrons, each ion makes the same contribution to τ^{-1} as f neutral molecules, where

$$f=\frac{6{\cdot}06\times10^{-3}}{1{\cdot}8\times10^{-8}}\left(\frac{300}{T}\right)^2=3{\cdot}37\times10^5\left(\frac{300}{T}\right)^2.$$

For example, the values of τ^{-1} quoted above for E, F_1 and F_2 can be explained, neglecting collisions with neutral molecules completely, by assuming the number-densities of ions in the three layers to be

$$10^7\left(\frac{T}{300}\right)^{\frac{3}{2}},\quad 6{\cdot}6\times10^5\left(\frac{T}{300}\right)^{\frac{3}{2}},\quad \text{and}\quad 3{\cdot}3\times10^5\left(\frac{T}{300}\right)^{\frac{3}{2}}/\text{cm.}^3.$$

These are of the same order as the usually assumed ion-densities.

However, only a small part of the collisions of electrons is likely to be with ions in the E-layer. During the day, the logarithm of the reflexion coefficient of wireless waves traversing the E-layer varies like $\cos^{\frac{3}{2}}\chi$, where χ is the sun's zenith angle; whereas, if collisions with ions were dominant, it would vary like $\cos\chi$. Hence the number of ions in the E-layer must be much less than $10^7(T/300)^{\frac{3}{2}}$/cm.^{3}. This may come about either because there actually are few ions or because T is very large (here T is the electron-temperature, not the temperature of the gas as a whole). Of these alternatives, the first appears the more probable. At any elastic collision with a molecule, a fast electron loses a fraction of its energy of order m_e/m_a, i.e. $(5\times10^4)^{-1}$. It undergoes 6×10^4 collisions per sec.; thus its energy is reduced in the ratio $1/e$ in $\frac{5}{6}$ sec. According to Massey & Bates (1943), the average free lifetime of an electron before attachment to a neutral atom is probably at least of order 30 sec. Hence, even though when first liberated its energy may be large, an electron is likely to possess an energy not much greater than the thermal throughout most of its free existence.

Since the gas temperature in the E-layer is not likely to differ greatly from 300° K, we conclude that the total number of ions must be much less than 10^7/cm.^{3}, i.e. much less than 50 ions of either sign per electron. To explain the observed collision-frequency, the number-density of neutral molecules must be about $3{\cdot}5\times10^{12}$/cm.^{3} at the centre of the E-layer. This corresponds to a pressure $\frac{1}{8}$ dyne/cm.^{2}, or 10^{-4} mm. Hg.

If Massey & Bates are correct in suggesting that there are few negative ions in F_1, only a relatively small fraction of the electron-collision in this layer can be with ions. Neglecting ionic collisions completely, and taking $T = 900°$ K, we find that

$n_a = 1{\cdot}3\times 10^{11}/\text{cm.}^3$. Massey & Bates suggest $n_a \sim 10^{11}/\text{cm.}^3$, using the absorption coefficient of atomic oxygen. For F_2, a large proportion of the electronic collisions may be with ions, and the collision-frequency does not, therefore, give any reliable information as to the neutral-gas density in this layer. The estimates for E and F_1 refer to England; at the equator, where ionizing radiation penetrates deeper, the densities may be rather greater—say

$$n_a = 4{\cdot}5\times 10^{12}/\text{cm.}^3 \text{ for E,} \quad \text{and} \quad n_a = 1{\cdot}6\times 10^{11}/\text{cm.}^3 \text{ for F.}$$

11. ELECTRON AND ION CONDUCTIVITIES

The electron conductivity (in e.m.u.) is given by

$$\begin{aligned} \sigma_e^{\mathrm{I}} + i\sigma_e^{\mathrm{II}} &= n_e^2 e^2/\{\theta_{ae} + \theta_{+e} + \theta_{-e} - iHn_e e\} \\ &= n_e e^2/\{m_e \tau^{-1} - iHe\} \\ &= n_e\{3{\cdot}57\times 10^{12}\tau^{-1} - iH\,.\,6{\cdot}29\times 10^{19}\}^{-1}. \end{aligned} \tag{58}$$

Thus σ_e^{II} and σ_e^{I} are equal if the collision-frequency τ^{-1} is $1{\cdot}76\times 10^7 H/\text{sec}$. This is at a height of about 85 km., where the total pressure is, say, 100 times that at the centre of the E-layer; here n_e and the conductivity are not likely to be large. In the E and F layers, σ_e^{I} is small compared with σ_e^{II}, and approximately

$$\sigma_e^{\mathrm{I}} = 9{\cdot}04\times 10^{-28} n_e H^{-2}\tau^{-1}, \quad \sigma_e^{\mathrm{II}} = 1{\cdot}59\times 10^{-20} n_e H^{-1}. \tag{59}$$

The maximum value of σ_e^{I} is reached rather below the centre of the E-layer; at midday at the equator it is of order $1{\cdot}5\times 10^{-16}$ e.m.u. In the F-layer, σ_e^{I} is probably never greater than 5×10^{-17} e.m.u. The transverse conductivity σ_e^{II} is larger; its maximum values in the E and F layers are about 10^{-14} and 10^{-13} e.m.u.

To obtain a general expression for the ion conductivity, the ternary-gas formula of § 7 must be used. Suppose that λ ($\equiv n_-/n_e$) is fairly large compared with unity; then the difference between n_+ and n_- can be neglected in calculating σ_i^{I}, but not in calculating σ_i^{II} (which vanishes if $n_+ = n_-$). If we put $n_+ = n_-$,

$$\begin{aligned} \sigma_i^{\mathrm{I}} &= \{\tfrac{1}{2}H^2\theta_{a+}^{-1} + (\theta_{+-} + \tfrac{1}{2}\theta_{a+})\, n_+^{-2} e^{-2}\}^{-1} \\ &= \left\{2{\cdot}22\times 10^{31}\frac{H^2}{n_a n_+} + 3{\cdot}32\times 10^{12}\left(\frac{300}{T}\right)^{\frac{3}{2}} + 4{\cdot}44\times 10^7\frac{n_a}{n_+}\right\}^{-1}. \end{aligned} \tag{60}$$

Numerical values of σ_i^{I} when $T = 300°$ K are illustrated in figure 2. When n_a/n_+ is greater than 10^6, the collisions of ions of opposite signs are unimportant; as a consequence, the lower curves in figure 2 are nearly identical in form. The flattening shown by the upper curves is the direct consequence of collisions of pairs of ions (represented by the second term in the bracket in (60)); because of such collisions, σ_i^{I} can never exceed $3\times 10^{-13}(T/300)^{\frac{3}{2}}$ e.m.u.

When the mutual collisions of ions can be ignored—when n_a/n_+ is greater than 10^6, or there are few negative ions—σ_i^{I} approximates to the sum of the conductivities due to the positive and negative ions separately, given by the uncorrected free-path formulae. Hence in this case

$$\begin{aligned}\sigma_i^{\mathrm{I}} + i\sigma_i^{\mathrm{II}} &= n_+^2 e^2(\theta_{a+} + in_+ eH)^{-1} + n_-^2 e^2(\theta_{a-} - in_- eH)^{-1} \\ &= n_+\{8{\cdot}89\times 10^7 n_a + i6{\cdot}29\times 10^{19}H\}^{-1} \\ &\quad + n_-\{8{\cdot}89\times 10^7 n_a - i6{\cdot}29\times 10^{19}H\}^{-1}. \qquad (61)\end{aligned}$$

This shows that σ_i^{II} is numerically less than σ_e^{II}, and opposite in sign to it. When n_a is less than $5\times 10^{11}/\mathrm{cm.}^3$, σ_i^{II} becomes comparable with σ_e^{II}, and it probably cancels most of σ_e^{II} in the F-layer.

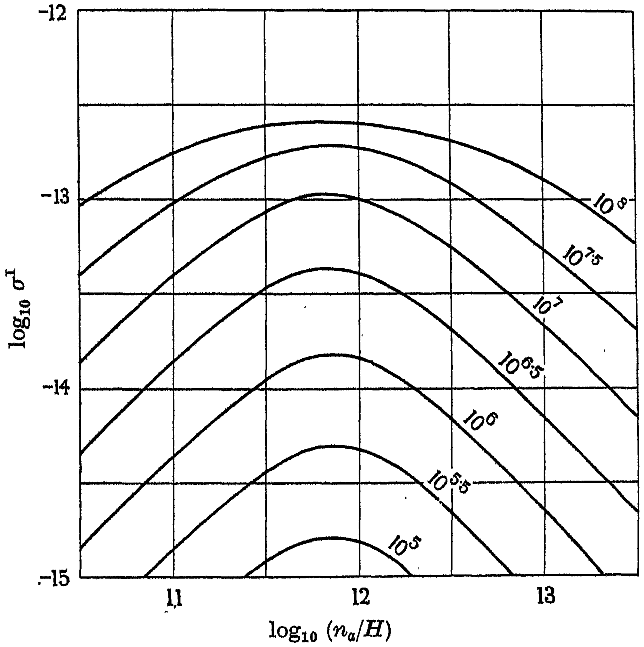


FIGURE 2. Graphs of $\log_{10}\sigma^{\mathrm{I}}$ against $\log_{10}(n_a/H)$ for certain values of n_+/H (shown against the corresponding curves).

Save at the base of the E-layer, σ_i^{I} is large compared with σ_e^{I}. Even if the number of negative ions is negligible, σ_i^{I} is 5×10^{-16} e.m.u. at midday at the equator at the centre of the E-layer, and with λ negative ions per electron ($\lambda \leqslant 50$) it is nearly $2\lambda + 1$ times as large. Were it not for the increase in dissociation of molecular oxygen with height, σ_i^{I} would take larger values throughout a thick layer above this level. According to the Chapman theory of layer formation, assuming no negative ions, n_e is roughly half its maximum value at a height at which n_a has decreased in the ratio $\frac{1}{10}$; this makes $\sigma_i^{\mathrm{I}} = 2\times 10^{-15}$ e.m.u. at a level close to the base of the F_1-layer. However, the E-layer arises from the ionisation of molecular oxygen; because of increasing dissociation of oxygen the conductivity of the E-layer is limited to a thickness which, for few negative ions, may be estimated as, say, 50 km. If a large number of negative ions of molecular oxygen are present, the thickness is likely to be smaller. For different values of λ, the integrated conductivity of the layer,

$$\int \sigma^{\mathrm{I}} dh \quad (h = \text{height}),$$

is estimated as shown

λ =	50	20	10	5	2	0
$\int \sigma^{\mathrm{I}} dh$ =	3×10^{-7}	$1{\cdot}4 \times 10^{-7}$	8×10^{-8}	$4{\cdot}5 \times 10^{-8}$	2×10^{-8}	5×10^{-9}

These estimates are, of course, very rough; they provide little more than an idea of the order of magnitude.

In the F_1-layer, the ion conductivity is appreciably affected by the magnetic field. Assuming no negative ions, σ_i^{I} is 8×10^{-15} e.m.u. at the centre of the layer, and similar in order of magnitude throughout much of the lower part of the layer; if there are an appreciable number of negative ions, σ_i^{I} is multiplied by $1 + 2\lambda$. At the centre of F_2, assuming $n_a = 4 \times 10^{10}/\text{cm.}^3$, it is found that $\sigma_i^{\mathrm{I}} = 1{\cdot}6 \times 10^{-14}$ e.m.u. Thus the integrated conductivities of the two layers, assuming widths 50 and 60 km. respectively, are 4×10^{-8} (at least) and 10^{-7} e.m.u.

It would appear from this that 'dynamo' currents due to tidal motion in the earth's atmosphere are not unlikely to flow mainly in the F-layer. This conclusion is, however, questionable, because of the drag due to induction. A motion across the lines of force, once started, tends to decay in a time $\rho(\sigma^{\mathrm{I}} H^2)^{-1}$ (cf. § 9); this is $45/(2\lambda + 1)$ days at the centre of the E-layer, $3\frac{1}{2}/(2\lambda + 1)$ hours at the centre of F_1, and only 20 min. at the centre of F_2.

Tidal motion may be compared with the forced harmonic motion given by the equation

$$\ddot{x} + \mu\dot{x} + \sigma^2 x = A e^{i\sigma' t},$$

the solution of which is

$$x = A e^{i\sigma' t}/(\sigma^2 - \sigma'^2 + \mu i \sigma').$$

In this, μ^{-1} is the above time of decay; the periods $2\pi/\sigma$, $2\pi/\sigma'$ of the free and forced oscillations are not very different (Pekeris 1937). Assuming that μ and σ are comparable, and $\sigma - \sigma'$ is small, the phase of the forced oscillation differs from that of the applied force by about a quarter of a period, and its amplitude is reduced, by the inclusion of the resistance term $\mu\dot{x}$, roughly in the ratio

$$(\sigma^2 - \sigma'^2)/\mu\sigma' = 2(\sigma - \sigma')/\mu.$$

Taking, for example, $2\pi/\sigma = 11$ hr., $2\pi/\sigma' = 12$ hr. (to get the correct order of magnitude) we find that tidal motions in F_1 are reduced in the ratio $1/3(2\lambda + 1)$, and those in F_2 in the ratio 1/33, by induction drag. The effect on the *magnitude* of the induced currents is roughly the same as if the tidal motion were unaltered, but the integral conductivities of the two layers were reduced to $1{\cdot}3 \times 10^{-8}$ and 3×10^{-9}

e.m.u.; the phase of the currents is altered roughly by a quarter of a period. By the same argument, an increase of λ above 50 in the E-layer, if such were possible, would not produce a great change in the magnitude of currents induced in this layer.

The arguments set out in § 10 suggest that λ cannot exceed (say) 25 in the E-layer. Thus the currents flowing in the E and F layers together are the same in amplitude as if induction drag were negligible, and the integral conductivity of the two layers taken together were between 2×10^{-7} and 2×10^{-8} e.m.u. The major part of this effective conductivity is contributed by the E-layer, unless $\lambda < 2$ in this layer. The currents are in phase with the tidal oscillation, or a quarter period out of phase, according as they flow chiefly in the E or F layer.

The effective conductivity thus obtained is smaller than earlier estimates; e.g. Chapman & Bartels (1940) suggested an integral conductivity of order 10^{-5} e.m.u. The chief reasons for the difference are the smaller width through which the E-layer is supposed to be conducting, and the smaller assumed value for λ in the E-layer. As seen above, even if λ could be much larger than assumed here, the electric currents in the E-layer would not be much larger, because of the increased effect on the conductivity of collisions between ions of opposite signs, and because of induction drag, neither of which have been considered earlier.

Chapman found that, to explain the lunar variation of the earth's magnetic field, the integral conductivity of the atmosphere at noon at the equator near sunspot maximum must be about $2{\cdot}5 \times 10^{-5}$ e.m.u. if the speed of the tidal currents is the same in the upper atmosphere as at the ground (Chapman 1919); if it is l times as great, the integral conductivity is $2{\cdot}5 \times 10^{-5} l^{-1}$ e.m.u. The difference between the conductivities parallel and transverse to the magnetic field was not realized at the time of Chapman's paper; since the resistance of a circuit is determined chiefly by the conductivity of its more resisting parts, the integral conductivity transverse to the field is probably of order $2 \times 10^{-5} l^{-1}$ e.m.u. Hence the estimates given above suggest that l is between 100 and 1000. This is not unreasonable, in view of the size of the tidal oscillation observed in the E-layer (Appleton & Weekes 1939).

The dynamo theory of the lunar variations in the earth's magnetic field is subject to the difficulty that the observed lunar tide in the E-layer appears to be opposite in phase to that required to explain the magnetic variations. The present work does not resolve this difficulty; but it may help to explain certain other minor phase-difficulties. Since the currents flowing in the E and F layers differ in phase, differences in phase between the solar and lunar variations in the earth's field may arise because of the different importance of the E and F layers in the two types of variation. Again, currents in the F-layer are chiefly determined by the balance between induction drag and tidal forces, and are nearly independent of the degree of ionization, while currents in the E-layer are roughly proportional to the ion-density; hence seasonal changes in phase are likely if the contributions of the two layers are comparable in importance.

12. DIAMAGNETISM AND DRIFT CURRENTS

Beside the dynamo theory of solar and lunar variations in the earth's field, other theories have also been examined, ascribing the variations either to the diamagnetic properties of ions and electrons, or to drift currents due to the combined action of gravitational, electric and magnetic fields. In another connexion (Cowling 1932) I have shown that the effects of diamagnetism and drift currents can also be described in terms of the electric currents arising from diffusion. These effects are accordingly discussed briefly here, even though Chapman & Bartels (1940) have set out weighty reasons why they are inadequate to explain the observed variations.

If λ is less than 25 in the E-layer, and small in the F-layer, the diffusion velocities of electrons, positive ions and negative ions are roughly independent. Using a formula like (47), and substituting for $\boldsymbol{d}_r$ from (21), the total electric current is found to be

$$\boldsymbol{j} = \sum_r (\theta_{ar}^2 + n_r^2 e_r^2 H^2)^{-1} \left\{ n_r e_r \theta_{ar} \left(-\frac{\partial p_r}{\partial \boldsymbol{r}} + \frac{\rho_r}{\rho} \frac{\partial p}{\partial \boldsymbol{r}} + n_r e_r \boldsymbol{E} \right) \right.$$
$$\left. - n_r^2 e_r^2 \boldsymbol{H} \wedge \left(-\frac{\partial p_r}{\partial \boldsymbol{r}} + \frac{\rho_r}{\rho} \frac{\partial p}{\partial \boldsymbol{r}} + n_r e_r \boldsymbol{E} \right) \right\}, \quad (62)$$

where the summation is over positive ions, negative ions, and electrons. In this, the terms involving $\boldsymbol{H} \wedge \partial p_r/\partial \boldsymbol{r}$ represent diamagnetism and a drift current due to inhomogeneity in the magnetic field, and the terms involving $\boldsymbol{H} \wedge \partial p/\partial \boldsymbol{r}$ (which is approximately $\boldsymbol{H} \wedge \rho \boldsymbol{g}$) and $\boldsymbol{H} \wedge \boldsymbol{E}$ represent the drift currents due to gravity and the electric field. It is necessary to retain the $\boldsymbol{E}$-terms; the relative diffusion of positive and negative charges produces a space-charge, whose field tends to oppose any further separation of positive and negative charges. Because of the horizontal stratification of the ionosphere, $\boldsymbol{E}$ will be vertical, as well as $\partial p_r/\partial \boldsymbol{r}$ and $\partial p/\partial \boldsymbol{r}$.

Here it must be recalled that (62) applies strictly only when the factors causing diffusion are transverse to the magnetic field, i.e. when $\boldsymbol{H}$ is horizontal. If $\boldsymbol{H}$ is not horizontal, (62) gives correctly only the component of $\boldsymbol{j}$ transverse to $\boldsymbol{H}$. Electron diffusion along the lines of force is unhampered by the magnetic field, and is much faster than ion diffusion. Thus, when $\boldsymbol{H}$ is not nearly horizontal, the electric field is due to electrons outstripping ions, and tends to reduce their diffusion velocity to equality with that of ions. That is, approximately,

$$n_e e \boldsymbol{E} = -\frac{\partial p_e}{\partial \boldsymbol{r}} + \frac{\rho_e}{\rho} \frac{\partial p}{\partial \boldsymbol{r}},$$

save in a narrow belt, one or two degrees of latitude in breadth, about the magnetic equator. This means that the electronic drift-current due to $\boldsymbol{E}$ nearly neutralizes that due to gravity and the currents due to electron diamagnetism, save in this belt. In this belt, however, the vertical diffusion of ions is faster than that of electrons; the electric field operates so as to increase the diffusion of electrons, and to amplify the electron diamagnetic and drift currents.

In the F-layer, the precise value of the electric force is unimportant. Since each θ_{ar} is small compared with $n_r e_r H$ here, the diamagnetic and drift currents are approximately

$$-\frac{1}{H^2}\sum_s \boldsymbol{H}\wedge\left(-\frac{\partial p_r}{\partial \mathbf{r}}+\frac{\rho_r}{\rho}\frac{\partial p}{\partial \mathbf{r}}+n_r e_r \boldsymbol{E}\right)=\frac{1}{H^2}\boldsymbol{H}\wedge\left(\frac{\partial}{\partial \mathbf{r}}(p_+ + p_- + p_e)-\frac{\rho_+ + \rho_- + \rho_e}{\rho}\frac{\partial p}{\partial \mathbf{r}}\right)$$

$$\doteqdot \frac{1}{H^2}\boldsymbol{H}\wedge\left(2\frac{\partial p_e}{\partial \mathbf{r}}-\frac{p_e}{p}\frac{\partial p}{\partial \mathbf{r}}\right), \tag{63}$$

since p_-, ρ_- and ρ_e are small, $p_+ \doteqdot p_e$ and $\rho_+/\rho = p_+/p$. To get an idea as to the order of magnitude of the total current, suppose that the variation of p_e with p (not necessarily with the height) is that for a simple Chapman layer. Then

$$p_e = P_e\left(\frac{p}{P}\right)^{\frac{1}{2}} e^{\frac{1}{2}(1-p/P)},$$

where P_e is the value of p_e at the level of maximum electron-density, and P is the corresponding value of p. Thus the currents (63) are

$$\frac{1}{H^2}\boldsymbol{H}\wedge\frac{P_e}{P}\left(\frac{p}{P}\right)^{\frac{1}{2}} e^{\frac{1}{2}(1-p/P)}\frac{\partial p}{\partial \mathbf{r}}.$$

Integrate with respect to height, neglecting the variation of $\boldsymbol{H}$ with height. Then if θ is the magnetic dip, the resulting integral current is of magnitude

$$H^{-1}\sin\theta\frac{P_e}{P}\int_0^\infty\left(\frac{p}{P}\right)^{\frac{1}{2}} e^{\frac{1}{2}(1-p/P)}\,dp=\sqrt{(2\pi e)}\,H^{-1}P_e\sin\theta = 4{\cdot}13H^{-1}P_e\sin\theta.$$

If the currents are supposed to flow in a thin current-sheet, the field they produce just below this sheet is

$$2\pi\times 4{\cdot}13H^{-1}P_e\sin\theta = 23H^{-1}P_e\sin\theta.$$

Taking $P_e = 2{\cdot}5\times10^{-7}$ dynes/cm.^{2} (corresponding to $1{\cdot}8\times10^6$ electrons/cm.^{3}, and a temperature 1000° K), this gives a field at noon near the equator at sunspot maximum of order 2×10^{-5} gauss $= 2\gamma$. To the order to which we have worked, the term of (63) involving $\partial p_e/\partial \mathbf{r}$ (the diamagnetic term) does not contribute to this field; the integral diamagnetic current is zero, and diamagnetism affects the field only because forward and reverse currents flow at different heights. The effect of diamagnetism on the field is therefore much smaller than 2γ.

In the E-layer, n_+ and n_- are probably at most of the same order as n_+ in F. In the E-layer, however, θ_{a+}/n_+ and θ_{a-}/n_- are large compared with eH; thus, in the absence of an electric force, ion currents are likely to be much smaller in E than in F. The same is true also of electronic currents, since n_e is much smaller in E than in F. The presence of an electric force does not greatly affect the situation. Except near the equator, the electric force increases the positive-ion currents, but decreases the negative-ion and electron currents by a rather larger amount. Near the equator the reverse is the case; if there are no negative ions, the electric field effectively transfers the forces causing diffusion from positive ions to electrons, and thereby increases the currents, since the drift of electrons is much less impeded by collisions than that of ions. If negative ions are present, the effect is much smaller, and in any case the currents in E cannot be more than comparable with those in F.

It is concluded therefore that diamagnetism and drift currents cannot produce a total field much greater than 2γ. The resulting diurnal variation of the field will be of order 2γ; semidiurnal changes must be much smaller. The inadequacy of the theory to explain the observed variations is therefore clear.

APPENDIX

Boltzmann's equation and ionized gases. At first sight it would appear that the Boltzmann equation does not apply to an ionized gas. This equation assumes that the motion of a molecule can be divided into relatively brief instants of encounter, separated by relatively long intervals during which forces of interaction with other molecules can be neglected in comparison with field forces. But in an ionized gas, because of the slow attenuation of electrostatic forces with increasing distance, interaction forces are everywhere at least comparable with field forces. For example, consider conditions in the reversing layer of the sun. The interaction force between two electrons r cm. apart is $2{\cdot}25 \times 10^{-19}/r^2$ dynes; the force on an electron moving with speed 5×10^7 cm./sec. (the mean thermal speed at 5750° K) in a magnetic field 1000 gauss is 8×10^{-10} dyne: these are equal only if $r = 1{\cdot}7 \times 10^{-5}$ cm., which is not much less than the mean distance between neighbouring charged particles. Most field forces are, moreover, very small compared with the above magnetic forces.

The position is saved by the fact that electrostatic forces, though large compared with field forces, are nevertheless not strong enough to produce large effects save at distances small compared with the mean distance between neighbouring charges. The mutual electrostatic energy of our two electrons is $2{\cdot}25 \times 10^{-19}/r$ erg; this equals the mean kinetic energy $2kT$ of relative motion only if $r = 1{\cdot}4 \times 10^{-7}$ cm. Thus, for every electrostatic encounter producing a large deflexion, there are many producing small deflexions; the cumulative effect of the latter outweighs that of the former. But, in an encounter producing a small deflexion, the effects of encounter and field forces are roughly additive; that is, the change in velocity can be represented by supposing that the molecules normally move under field forces alone, but that at a given instant their velocities are suddenly changed by the full effect of the interaction forces. The motion then becomes of the type to which Boltzmann's equation can be applied.

It may appear surprising that, even though magnetic forces on a charge are not large compared with interaction forces, they dominate the interaction forces in a sunspot to the extent that conditions of 'free spiralling' are approached. The reason again is that a single encounter rarely produces a sizeable deflexion, and successive encounters produce deflexions in different directions. Thus the magnetic field, acting consistently in the same sense, is able to build up a large deflexion sooner than the random efforts of encounters.

In conclusion, I should like to thank Professor S. Chapman for his continued interest and helpful advice during the progress of this work.

REFERENCES

Appleton, E. V. & Weekes, K. 1939 *Proc. Roy. Soc.* A, **171**, 171.
Chapman, S. 1919 *Phil. Trans.* A, **218**, 1.
Chapman, S. & Bartels, J. 1940 *Geomagnetism.* Oxford Univ. Press.
Chapman, S. & Cowling, T. G. 1939 *The mathematical theory of non-uniform gases.* Camb. Univ. Press.
Cowling, T. G. 1932 *Mon. Not. R. Astr. Soc.* **92**, 407.
Massey, H. S. W. & Bates, D. R. 1943 *Rep. Prog. Phys.* Physical Soc. Lond. **9**, 62.
Pekeris, C. L. 1937 *Proc. Roy. Soc.* A, **158**, 650.
Townsend, J. S. & Tizard, H. T. 1913 *Proc. Roy. Soc.* A, **88**, 336.
Tyndall, A. M. 1938 *Positive Ions.* Camb. Univ. Press.

INDEX TO VOLUME 183 (A)

Absorption of radium (B+C) gamma-rays (Roberts), 338.
Acetone, thermal decomposition (Smith & Hinshelwood), 33.
Allen, D. N. de G. & Southwell, R. V. Relaxation methods applied to engineering problems, X, 125.
Allen, D. N. de G., Southwell, R. V. & Vaisey, Gillian. Relaxation methods applied to engineering problems, XI, 258.
Anniversary address (Dale), 243.

Balmer spectrum, long duration in excited hydrogen (Rayleigh), 26.
Bell, R. P. The occurrence and properties of molecular vibrations with $V(x) = ax^4$, 328.
Bell, R. P. & Longuet-Higgins, H. C. The normal vibrations of bridged X_2Y_6 molecules, 357.
Bhabha, H. J. & Harish-Chandra. On the theory of point-particles, 134.
Bruce, C. E. R. The initiation of long electrical discharges, 228.

Chadwick, J., May, A. N., Pickavance, T. G. & Powell, C. F. An investigation of the scattering of high-energy particles from the cyclotron by the photographic method, I. The experimental method, 1.
Champion, F. C. & Powell, C. F. Application of the photographic method to problems in nuclear physics, II. The scattering of 8·8 and 13 MeV neutrons by protons, 64.
Chang, T. S. A note on the Hamiltonian theory of quantization, 316.
Continuous spectra in flames, role of atomic oxygen (Gaydon), 111.
Cosmic-ray bursts, production by mesotrons (Mohr & Stafford), 54.
Cowling, T. G. The electrical conductivity of an ionized gas in a magnetic field, with applications to the solar atmosphere and the ionosphere, 453.
Cyclotron, scattering of high-energy particles (Chadwick & others), 1.

Dale, Sir Henry. Anniversary address, 243.
Daniels, H. E. The statistical theory of the strength of bundles of threads, I, 405.
Dirac, P. A. M. Unitary representation of the Lorentz group, 284.
Dissociation energy of CO, N_2, NO and CN (Gaydon & Penney), 374.

Electrical conductivity of an ionized gas in a magnetic field (Cowling), 453.
Electromagnetic fields, electron swarms (Gabor), 436.
Equation of state for solids (Fürth), 87.

Fibrous proteins, combination with acids (Gilbert), 167.
Flow of gas through a nozzle with velocities exceeding the speed of sound (Fox & Southwell), 38.
Fox, L., & Southwell, R. V. On the flow of gas through a nozzle with velocities exceeding the speed of sound, 38.
Fürth, R. On the equation of state for solids, 87.

Gabor, D. Stationary electron swarms in electromagnetic fields, 436.
Gaydon, A. G. Continuous spectra in flames: the role of atomic oxygen in combustion, 111.

Gaydon, A. G. & Penney, W. G. The dissociation energies of CO, N_2, NO and CN, 374.
Gilbert, G. A. The combination of fibrous proteins with acids, II, 167.
Guggenheim, E. A. Statistical thermodynamics of mixtures with zero energies of mixing, 203.
Guggenheim, E. A. Statistical thermodynamics of mixtures with non-zero energies of mixing, 213.

Hamiltonian theory of quantization (Chang), 316.
Harish-Chandra. On the removal of the infinite self-energies of point-particles, 142.
Harish-Chandra. *See also* Bhabha & Harish-Chandra.
Hinshelwood, C. N. *See* Smith & Hinshelwood.

Jánossy, L. Some aspects of the production of mesons and the barometer effect of penetrating showers, 190.
Jánossy, L. & Rochester, G. D. The barometer effect of penetrating showers, 186.
Jánossy, L. & Rochester, G. D. The transition effect of penetrating showers, 181.
Jeffrey, G. A. The structure of polyisoprenes, I, 388.

Long electrical discharges, initiation (Bruce), 228.
Longuet-Higgins, H. C. *See* Bell & Longuet-Higgins.
Lorentz group, unitary representation (Dirac), 284.

Magnetic field, electrical conductivity of ionized gas (Cowling), 453.
May, A. N. *See* Chadwick & others.
Mohr, C. B. O. & Stafford, G. H. The production of cosmic-ray bursts by mesotrons, 54.
Molecular vibrations with $V(x) = ax^4$ (Bell), 328.

Normal vibrations of bridged X_2Y_6 molecules (Bell & Longuet-Higgins), 357.

Penetrating showers, barometer effect (Jánossy), 190.
Penetrating showers, barometer effect (Jánossy & Rochester), 186.
Penetrating showers, transition effect (Jánossy & Rochester), 181.
Penney, W. G. *See* Gaydon & Penney.
Photographic method, applications to nuclear physics, II (Champion & Powell), 64.
Pickavance, T. G. *See* Chadwick & others.
Point-particles, removal of infinite self-energies (Harish-Chandra), 142.
Point-particles, theory (Bhabha & Harish-Chandra), 134.
Polyisoprenes, structure (Jeffrey), 388.
Powell, C. F. *See* Chadwick & others.
Powell, C. F. *See* Champion & Powell.

Rayleigh disk, performance (Scott), 296.
Rayleigh, Lord. Long duration of the Balmer spectrum in excited hydrogen, 26.
Relaxation methods applied to engineering problems, X (Allen & Southwell), 125.
Relaxation methods applied to engineering problems, XI (Allen, Southwell & Vaisey), 258.
Roberts, J. E. Absorption of radium (B+C) gamma-rays, 338.
Rochester, G. D. *See* Jánossy & Rochester.

Scott, R. A. An investigation of the performance of the Rayleigh disk, 296.
Smith, J. R. E. & Hinshelwood, C. N. The thermal decomposition of acetone, 33.
Southwell, R. V. *See* Allen & Southwell.

Southwell, R. V. *See also* Allen, Southwell & Vaisey.
Southwell, R. V. *See* Fox & Southwell.
Stafford, G. H. *See* Mohr & Stafford.
Stationary electron swarms in electromagnetic fields (Gabor), 436.
Statistical theory of the strength of bundles of threads, I (Daniels), 405.
Statistical thermodynamics of mixtures with non-zero energies of mixing (Guggenheim), 213.
Statistical thermodynamics of mixtures with zero energies of mixing (Guggenheim), 203.

Thermal decomposition of acetone (Smith & Hinshelwood), 33.

Vaisey, Gillian. *See* Allen, Southwell & Vaisey.

END OF THE ONE HUNDRED AND EIGHTY-THIRD VOLUME (SERIES A)

I. A. R I. 75.

INDIAN AGRICULTURAL RESEARCH
INSTITUTE LIBRARY
NEW DELHI.

Date of issue.	Date of issue.	Date of issue
....	....	....
....	....	....
....	....	....
....	....	....
....	....	....
....	....	....
....	....	....
....	....	....
....	....	....
....	....	....
....	....	....
....	....	....
....	....	....
....	....	....

S. C. P.—1/8/47 P. J.-3-5-48-2000